

COK-4103335-13- 0335087

AMERICAN PHILOSOPHICAL QUARTERLY

13

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Myles Brand
Robert Butts
Roderick M. Chisholm
L. Jonathan Cohen

James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Carl G. Hempel
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Richard Rorty
Jerome Schneewind
George A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 13/NUMBER 1

JANUARY, 1976

CONTENTS

- | | | | |
|---|----|---|----|
| I. H. J. McCLOSKEY: <i>Human Needs, Rights and Political Values</i> | I | VII. J. B. MAUND: <i>The Non-Sensuous Epistemic Account of Perception</i> | 57 |
| II. RONALD D. MILO: <i>The Notion of a Practical Inference</i> | 13 | VIII. STEWART CANDLISH: <i>The Incompatibility of Perception: A Contemporary Orthodoxy</i> | 63 |
| III. R. H. KANE: <i>Nature, Plenitude and Sufficient Reason</i> | 23 | IX. EIKE VON SAVIGNY: <i>Listener-Oriented versus Speaker-Oriented Analysis of Conventional Meaning</i> | 69 |
| IV. FRANK JACKSON: <i>The Existence of Mental Objects</i> | 33 | X. W. D. GLASGOW: <i>Psychological Egoism</i> | 75 |
| V. FRANK B. DILLEY: <i>The Status of Religious Beliefs</i> | 41 | XI. JAMES F. HARRIS, JR.: <i>Part-of-the-Meaning-of-a-Word</i> | 81 |
| VI. JOSÉ FERRATER-MORA: <i>On Practice</i> | 49 | XII. T. S. CHAMPLIN: <i>Punishment Without Offence</i> | 85 |

© AMERICAN PHILOSOPHICAL QUARTERLY, 1976

PUBLISHED IN OXFORD BY BASIL BLACKWELL

ISSN 0003-0481

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by [✓] ~~authors~~ of any country on any aspect of philosophy, substantive or historical. However, only self-sustaining [✓] ~~articles~~ will be published, and not news items, book reviews, critical notices, or "discussion notes".

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The price *per annum* is twelve dollars for individual subscribers and twenty-eight dollars for institutions. Checks and money orders should be made payable to the publishers, Basil Blackwell & Mott Ltd. Back issues are sold at the rate of three dollars to individuals, six dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, U.K.).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The back cover of the journal may be consulted for details.

335087



AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Alvin Plantinga
Jerome Schneewind
George. A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton



VOLUME 13 (1976)

PUBLISHED IN OXFORD BY BASIL BLACKWELL

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 13 (1976)

ALSTON, WILLIAM P.: <i>Self-Warrant: A Neglected form of Privileged Access</i>	257
AXINN, SIDNEY AND DAVID AXINN: <i>Notes on the Logic of the Ignorance Relations</i>	135
BAIER, ANNETTE: <i>Mixing Memory and Desire</i>	213
BARKER, JOHN A.: <i>What You Don't Know Won't Hurt You?</i>	303
CANDLISH, STEWART: <i>The Incompatibility of Perception: A Contemporary Orthodoxy</i>	63
CHAMPLIN, T. S.: <i>Punishment Without Offence</i>	85
COBURN, ROBERT: <i>The Persistence of Bodies</i>	173
CORRIGENDA	315
DILLEY, FRANK B.: <i>The Status of Religious Beliefs</i>	41
FERRATER-MORA, JOSÉ: <i>On Practice</i>	49
FLYNN, J. R.: <i>The Realm of the Moral</i>	273
FRYE, MARILYN: <i>On Saying</i>	123
GLASGOW, W. D.: <i>Psychological Egoism</i>	75
HARRIS, JAMES F., JR.: <i>Part-of-the-Meaning-of-a-Word</i>	81
HARRIS, ZELLIG: <i>A Theory of Language Structure</i>	237
HEATH, PETER: <i>The Incredulous Hume</i>	159
JACKSON, FRANK: <i>The Existence of Mental Objects</i>	33
KANE, R. H.: <i>Nature, Plenitude and Sufficient Reason</i>	23
LESLIE, JOHN: <i>The Value of Time</i>	109
LEWIS, DAVID: <i>The Paradoxes of Time Travel</i>	145
MASSEY, GERALD J.: <i>Tom, Dick, and Harry, and All the King's Men</i>	89
MAUND, J. B.: <i>The Non-Sensuous Epistemic Account of Perception</i>	57
McCLOSKEY, H. J.: <i>Human Needs, Rights and Political Values</i>	1
MICHAEL, FREDERICK SEYMOUR: <i>What is the Master Argument of Diodorus Cronus?</i>	229
MILLER, CLYDE LEE: <i>Ironie or Not?</i>	309
MILO, RONALD D.: <i>The Notion of a Practical Inference</i>	13
OAKLEY, I. T.: <i>An Argument for Scepticism Concerning Justified Beliefs</i>	221
ODEGARD, DOUGLAS: <i>Parasitical Reference and Paradox</i>	295
OLIN, DORIS: <i>Newcomb's Problem: Further Investigations</i>	129
SAVIGNY, EIKE VON: <i>Listener-Oriented Versus Speaker-Oriented Analysis of Conventional Meaning</i>	69
SCHIFFER, STEPHEN: <i>A Paradox of Desire</i>	195
SCHLESINGER, GEORGE: <i>An Important Necessary Difference Between People and Mindless Machines</i>	205
SUTER, RONALD: <i>The Dream Argument</i>	185
URMEON, J. O.: <i>Fiction</i>	153
WARD, ANDREW: <i>Direct and Indirect Realism</i>	287
WÄSSERSTROM, RICHARD: <i>The University and the Case for Preferential Treatment</i>	165

I. HUMAN NEEDS, RIGHTS AND POLITICAL VALUES

H. J. McCLOSKEY

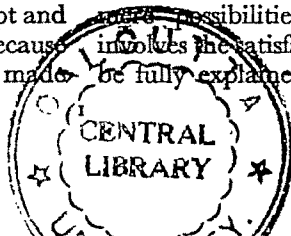
IN this paper I am concerned to raise for consideration the question of the nature of the connexion, if any, between human needs and human rights, individual needs and individual rights, human and individual needs and political values such as justice, equality, freedom, self-development, fraternity and privacy, and, more generally, to canvass the relevance of human and individual needs in the context of political philosophy generally. It would seem obvious that the fact that human beings have needs such as for food, drink, shelter, sleep, affection, the society of and communication with other persons, and the like, is relevant data to political theory. How, and in what way, is its relevance to be noted in political theory?

It is notable that reference to needs is seldom or sparingly explicitly made in those terms in the writings of the great philosophers of liberalism, Locke, Humboldt, Mill, Green, Spencer, and Hobhouse, to mention only some such philosophers, in the writings of whom one might reasonably have expected significant attention to the nature of human needs. In the case of J. S. Mill, where others might appeal to needs, we find instead, reference to human and individual goods and interests, the notion of interests being a key one in his philosophy. With other liberal philosophers, it is the individual's good rather than his needs, that is appealed to. Given the importance attached to needs, and the very considerable importance ascribed to the distinction between real and false needs by so many today, it would seem to be worthwhile to examine the concept of a need to determine whether the seeming minor role, indeed, the seeming complete neglect explicitly to consider human needs by so many notable political philosophers, including those who have contributed to the political philosophy of our liberal society is really what it appears to be, whether something important has been neglected or whether what is important in respect of human rights and political values can be determined with little or no reference to the concept and content of our needs. This is important also because there have been philosophers who have made

something of needs. It is fashionable today to stress this aspect of Marx's writings, and Marcuse's appeal to and use of theories about needs are very well known.

Again, there are many who believe that the logical gap between "is" and "ought" statements can be bridged by need statements, that we can move in a logically impeccable way from statements of fact to statements about what we or the state ought to do, where the statements of fact are statements about needs. Others seem to believe that need statements are simply another way of expressing statements about rights, and hence, that we can settle questions about rights with comparative ease by way of factual investigations into the nature of our needs. Others take a more pessimistic view, and treat need statements as simply another way of expressing value and ought statements, as statements which cannot be verified or vindicated simply by way of empirical inquiry, and as statements which depend for their defence on the defence of certain normative judgments about what is desirable or what ought to be done or brought about. Others again see need statements as simply socially or culturally relative statements or claims as to what are reasonable expectations or demands in particular societies, with no claim to any general or inter-society, inter-culture validity. In view of this extreme range of views about the relevance and importance of need statements to statements about rights and political values, it seems to be important to examine the import and significance of need statements.

However, the aims of this paper are modest and tentative. I am concerned to suggest that many today overrate the importance of needs theory as a part of political theory; that human needs can usefully be explained and appealed to only against a background belief in the existence of a common human nature and specific individual natures with certain inherent potentialities (as distinct from mere possibilities), the development of which involves the satisfaction of needs; that rights cannot be fully explained simply in terms of needs, and



that a rights theory which confined its claims to the contention that we have a *prima facie*, abstract, ideal right to the satisfaction of our needs, and only to this, would be claiming too few rights, and at the same time, the impossible and the absurd. In respect of the latter, it may be observed here that we may have a need for affection; it is not clear that any sense can be given to a claim that we have a right, even a *prima facie* negative right to affection. On the other hand, we have a right to much more by way of shelter, food, drink, etc., than would be dictated simply by what is necessary for the satisfaction of our needs. Further, the discovery of our needs does not help us to bridge the is-ought gap, to move from fact to value, for the discovery of *human* needs itself involves value judgments, that our existence and development are goods or means to or conditions of goods, or of lessening evils of human beings. Those who deny that there is some common human nature, and particular natures with inherent potentialities for good development, on the model of the acorn and the oak, or in terms of the teleological model of human nature of Thomistic natural law theory, would do better to drop talk about needs. Instead, I suggest, that they would be better advised to concentrate on the idea of the possibilities of changes in man and men, for the sake of realizing goods, reducing and eliminating evils, for, for such theories, it is the possibility of future goods which may be achieved by what, following Bradley and others, might usefully be called self-realization or self-making, rather than by self-development or self-perfection that is important.

"Need" the verb and "need" the noun: Many philosophers, writing of needs, move from statements about what is needed, to conclusions about what are needs, and what are human needs.¹ I suggest that this kind of approach can be seriously misleading, and that the noun "need" must be distinguished from the verb "to need," and that the differences in the context of "human needs" are even more marked than with other animate objects. Further, whilst we can speak of what inanimate objects need, we cannot except metaphorically or analogically speak of their needs. The soup may need salt and parsley but it does not have a need for salt and parsley. I may need a

ticket to go to the Grand Final, but unless in some real way my well-being as a person depends on my attending the Grand Final, I do not have a need for a ticket for the Grand Final. Similarly, as a footballer, Smith may need boots which allow various arrangements of stops to be used according to the state of the ground, but he cannot be said to have a need for such boots, although *qua* footballer (and not *qua* the person Smith) he may need them, and even have a need for them, but as a person, he would have a need for them only if his life or well-being in certain specific respects depended on his being a footballer and his playing well. That his winning the medal for the best and fairest player depended on his having such boots, would not in itself be sufficient to entitle us to say that he, Smith, has a need for them. Rather we should speak of his *interests*, and say that it was in his interests as the person Smith in so far as his interests as a footballer contributed to his interests as a person. If he played amateur football with his local church team of a Saturday morning, it is unlikely we should talk about his interests here. It is true that we can and do properly speak of needs which go with certain stations or roles in life, and, in these contexts, the use of the verb or noun may be optional. However, the statement using the verb is always incomplete. We need always to be told for what the thing needed, is needed. Talk about what individuals need, is commonly context-dependent, and is often simply a brief way of indicating what it is necessary for them to have to perform some task, satisfactorily occupy some role or station, or successfully to realize some end. Hence statements of the kind "Jones (who is a judge) needs a wig," cannot without error or misleadingness be abbreviated into "Jones has a need, namely, for a wig," "Among Jones' needs are shelter, clothing, X number of calories a day, and a wig." Need (noun) statements are also in a sense incomplete, but they are not context-dependent in the same way on the kind of reason that may be advanced to support a claim that X has a need. White is correct when he argues of need (verb and noun talk) that it is elliptical, but with talk about needs (noun) the range of reasons that properly can be given is much more limited; and of human needs more limited again than with need (verb) statements. White is also right that the elliptical nature of such talk has

¹ There is some evidence of this in the discussions of Brian Barry, *Political Argument* (London, 1965), pp. 47-49, S. I. Benn and R. S. Peters, *Social Principles and the Democratic State* (London, 1959), ch. 6, sn. III, Joel Feinberg, *Social Philosophy* (Englewood Cliffs, 1973), pp. 111-112, and in a more subtle, qualified way, by Alan R. White, "Needs and Wants" in *Philosophy Of Education Society Proceedings*, 1974, pp. 159-166, esp. 159-160.

been confused with its involving a value judgment. Need talk as he understands it need not do so; nor does talk about the needs of organisms other than man involve such value judgments. However, as against White, it will be argued in this paper that talk of *human needs* does involve value judgments. There is a further important difference between need talk as noted by White and talk of human needs, which is what is of importance to political and social theory, namely that in the latter case it is necessary to distinguish needs from *luxuries*, whereas in the contexts in which the verb is used, we may quite properly speak of needing something which is a luxury, a gross extravagance, for example, a second unlisted phone for our second best Rolls Royce. (See "Needs and Wants," *op. cit.*, pp. 160-161.)

It will be useful to enlarge upon and further support various of these contentions by contrasting them with those advanced by White. White approaches "need talk" along the lines of observing that inanimate as well as animate things can need; that among things that are needed are what should ordinarily be called luxuries; that for every statement about needing *X* or to *V*, there is a corresponding statement about a need or the need. White's examples of need statements include: an engine may need more oil; a valid syllogism needs at least one universal premiss; curtains may need to be lengthened; a student may need to work harder; a young woman well-equipped with shoes may need shoes of the latest fashion; a person needs to be at least 70 to qualify for the pension. It is true that we use the verb in these ways, in ways which would allow us to use instead the locution "it is necessary." However, we are not necessarily talking of needs (noun) when we speak of what is necessary for a valid syllogism, an efficiently running engine, gaining a 1st, keeping up with the fashions, qualifying for a pension. White suggests the contrary, arguing that for every need (verb) statement, there is a corresponding need (noun) statement: "Whenever *A* needs *X* or needs to *V*, there is also a or the, need for *A* to have *X* or to *V*; and *vice versa*. But such a need is not necessarily *A*'s need; and, therefore, although if *A*'s need is for *X* or to *V*, then *A* needs *X* or to *V*, the converse is not necessarily true. In other words, to have a need implies to need, whereas to need does not necessarily imply to have a need, though it does imply that there is a need" (*ibid.*, pp. 159-160).

Presumably the need (noun) statements about the syllogism and the engine would be: "There is a

need for a valid syllogism to have a universal premiss," "There is a need for engines of this type to have oil." I suggest that this use of "need" is secondary, parasitic on the primary use, according to which we speak of the needs of plants and animals, for water and food. Syllogisms and engines do not have needs for anything—White implicitly seems to acknowledge this. The locution in terms of "There is a need" is awkward, and is logically quite distinct from talk of needs, where the needs have a possessor. That there is an important difference is confirmed by the fact that whilst we may conceivably allow that it might be said that there is a need for shoes of the latest fashion, a second phone in the Rolls Royce, if certain conditions are to be met, purposes to be achieved, we should not allow that the young woman has a need for the fashion shoes, the multimillionaire the second phone, without a vast amount of supporting evidence relating to how these bear vitally on them as human persons, or as the specific individuals they are. Rather, we should object that the demands for the shoes, the phone, are demands not for needs but for luxuries. Any account of needs which allows luxuries to count as needs, must surely be mistaken, for this is a basic contrast in the context of talk about needs. Further, as I shall argue later, in the context of human needs, although not that of talk of needs of plants and animals, for example, of weeds, parasites, disease organisms, there is a value judgment involved, such that the judgments about human needs are not simply elliptical, which makes for this talk of human needs being even more importantly different from talk about need as expressed by statements using the verb, to need.

To consider now this more basic, seemingly much more important use of the noun "need," according to which we speak of a person's needs as a man, or as that particular person as that person. It is true that here we also commonly use the verb to point to such needs—we talk of particular persons needing this or that, or of human beings needing certain things—it is useful to confine the use of the noun to those needs that we have as men, or as the persons we are. Human needs, the needs of men as men, and the individual needs of particular persons as the unique persons they are (and not as holders of roles and stations) are the needs with which we are most concerned, when we are concerned with *human and personal rights* and political values. However, it is of course true that certain social, quasi-natural roles such as those of

mother, father, parent, which become basic to the fabric of most human social organizations, as well as some which are basic only to certain social organizations, may also be of importance when considering specific needs, rights, values to be realized, but these are secondary to the basic needs of men as men, and those of individual persons as those unique persons.

Human needs: It is important, at the conceptual level at least, to distinguish *natural drives, desires, wants, expectations, reasonable expectations, interests, welfare, well-being, luxuries*, from *needs*. *Drives*, including *natural drives* (which may be conscious or unconscious) could be destructive. For example, we could have a death drive. Hence it is a matter of empirical inquiry to determine what are our natural and acquired drives, and whether they relate to needs, or whether they lead to some kind of lack of fulfilment, a frustration, a destruction of the self. This is most evidently true of the drives which occur in particular individuals. If the drive becomes strong enough in a particular person, then, whether it is basically a natural or an acquired one, we may wish to describe it as a compulsion because we see it as something to be checked and even eliminated, if this is possible, rather than as something to be appeased as relating to a need. As a contingent fact about human nature, there seem in fact to be no such destructive human drives, but this is simply a contingent fact, and further research may upset this belief.

Natural desires, as well as *socially inculcated and acquired desires* equally ought at the conceptual level to be distinguished from needs. Butler showed awareness of this in respect of hunger when he observed that it was only because of the benevolence of God that, when we felt hunger, we desired food, not glass and stones. We could, perhaps only with a different physical constitution, have been made to hunger for glass and stones and not food, and hence to die of internal injuries and/or starvation. Such a desire would not have related to a need. Similarly, suppose that what certain experts claim to be true is true, that human beings are bi-sexual, having desires for sexual experiences with members of both sexes. This would not establish that they have a need for such experiences. So too, to establish that human beings by virtue of their natures have sexual desires, is to establish very much less than that they have needs relating to these desires. A possible step in the latter direction

would consist in showing, as exponents of the Thomistic natural law have sought to show, that the desire is teleological, having as an end, the realization of the continuance of the human race, this being related to perfection of men as men. Compare here the generally acknowledged desires for self-preservation and pleasure. Whilst we may need to preserve ourselves to attain a desired goal, and to experience pleasure in order to be happy, in neither case is talk about corresponding needs to these desires evident nor evidently appropriate. I suggest that human beings do not have a need for life nor for pleasure, although it can reasonably be argued that we have needs for what is essential for the maintenance of life, and perhaps for pleasure and the absence of pain. Nonetheless, it is true that there is often a connexion between natural desires and needs—perhaps for the reason Butler indicated, that we have a benevolent Creator, or because of “the survival of the fittest,” or simply due to chance.

Talk about *wants* is typically imprecise. What is called a want may be a drive, a desire, a wish, or even what is called by some, a felt need. (See Jonathan Bradshaw in *New Society*, Vol. 19 (1972), p. 641.) Again, to have a want is clearly distinct from having a need. What the relation between a want and a need is, if there is one at all, obviously depends on what is meant by “want,” and on the nature of the want. In his discussion of wants, White in “Needs and Wants,” whilst noting that neither need nor want imply a lack, and are distinct from a lack, draws a number of distinctions, some of them important, between wants in a sense akin to wishes, and needs. Some distinctions he seeks to draw depend on his claims that inanimate objects can need, and that need (verb) statements relating to them can be translated into or imply need (noun) statements. The latter claim was rejected earlier in this paper. White would seem to be right in noting that statements about wants in this sense of want are not necessarily elliptical, that the idea of a reason is differently related to needs and wants, that there is a connexion between wants and beliefs which does not hold of needs, and that whilst it is possible, although strange and unusual not to know what you want, this is not so with needs, for with the latter we have to inquire to determine what they are. Obviously we are more likely to be mistaken about our needs than about our wants.

To consider now needs, and what have sometimes been identified with or related to them,

namely, *expectations* and *reasonable expectations*. Although Benn and Peters open their analysis of need by observing that when we attribute to anyone a need: "It is, first, to indicate the lack of something which it would be injurious or detrimental to the subject not to supply; alternatively, a lack which frustrates some end envisaged on his account" (*op. cit.* p. 143), they go on to explain basic needs in terms of conceptions of living in a certain manner, arguing that: "They (statements concerning basic needs) suggest what seems the bare minimum for a 'decent' sort of life; and this varies from time to time and place to place" (*ibid.*, pp. 144-145). And: "'Basic needs' are thus a function of the general living standards of the community in question which yield norms like 'subsistence level', or 'a decent standard of living'" (p. 145). And: "Basic needs are thus fairly precisely determinate because they are related to norms set by conditions already very widely enjoyed. They are needs for precisely those things that most people have got. Beyond this minimum, however, needs become increasingly vague" (p. 145). The kinds of conclusions reached by J. F. Cronin concerning what is a just wage in terms of "the basic requirement for all men to live as befits normal human beings" and not solely by reference to needs, leads as he notes to what is a luxury today becoming a conventional necessity tomorrow.² Others, less careful than Cronin, have in this context moved towards explaining needs as expectations or as reasonable expectations, thereby acknowledging that on their view needs vary from society to society, and that they therefore are not related simply to men as men, but to men as members of specific societies, or as holders of certain roles in those societies. I wish to reject the kind of social and cultural relativity involved in such accounts of needs. Those who so argue may well be right about the sociology of talk about needs or basic needs—what are claimed to be needs in any particular society may well be what is accepted as "decent," "normal," "reasonable" in that society. However, such claims can be mistaken in ways not allowed for in such accounts. The needs claimed can be excessive or inadequate, and the ways in which they are to be seen to be such is not by looking into what are actual or reasonable or normal expectations, but by considering human nature, human potentialities *for good*, and what is

necessary for the continued existence of the possessor of that nature in a way that is not impaired or marred, and for the actualization of these potentialities for good. To illustrate: In keeping with the spirit if not the letter of the account of basic needs of Benn and Peters and of those who argue in like vein, is the conclusion that the better cared for slaves of Athens and pre-Civil War America had their needs fully met, whilst the impoverished Californian who cannot afford a house with sufficient rooms to allow each child a room of its own, like the family that cannot afford a colour TV set, and with a house the size of which necessitates the children of the same sex sharing a room, do not have their needs fully met (other factors being assumed to hold). This conclusion obviously is untenable. This is so because needs relate to natural goods, to goods determined by the individual's nature, and not simply by what is usual, normal, or expected in his society.

The initial suggestion that a need is something which it is injurious or detrimental not to supply, if, but only if, the detrimental is explained in terms of the individual's or human nature viewed as good, is nearer the mark. It is similar to that more recently set out by Joel Feinberg in *Social Philosophy* thus:

"The concept of a 'need' is extremely elastic. In a general sense, to say that *S* needs *X* is to say simply that if he doesn't have *X* he will be harmed. A 'basic need' would then be for an *X* in whose absence a person would be harmed in some crucial and fundamental way, such as suffering injury, malnutrition, illness, madness, or premature death. Thus we all have a basic need for foodstuffs of a certain quantity and variety, fuel to heat our dwellings, a roof over our heads, clothing to keep us warm, and so on." (p. 111.)

However, these suggestions too need to be modified. The suggestion that we have a need for that, the absence of which will involve us in harm, including injury, is too inclusive, even though by his examples Feinberg shows that he does not mean *any harm*, as to one's finances, career, but to one's person. Whilst it is clear that most of us will harm ourselves or come to harm as a result of our diets and lack of proper exercise, it is not clear that we have a need for health policemen to prevent this harm. Similarly, it might well be asked, Do we have a need for safety belts in our cars? Does the child that cannot feel pain, have a need for pain sensitivity

² J. F. Cronin, *Catholic Social Principles* (Milwaukee, 1950), p. 353. Cronin also writes: "The standard of a decent living wage varies according to time and place. It is relevant to the resources and economic organization of a given country." (p. 354.)

in so far as it is exposed to harm because of the lack of it? Do we have a *need* for all that knowledge, the absence of which exposes us to harm? I suggest that harm is too encompassing a word, injury too positive and yet too narrow, and that needs relate to what it would be detrimental to us to lack, where the detrimental is explained by reference to our natures as men and specific persons. Thus it would be detrimental to us as human beings to lack food, drink, clean air; the lack of these things would be harmful to our health and hence normally to our well-being as persons, even though the lack could hardly be described as involving an injuring of us, unless it was due to a deliberate withdrawal of them by another person, and even then, we might well be reluctant to describe the theft or non-provision of an earned food as constituting the doing of an injury to us. I suggest that the lack, withholding, not supplying of what is dictated by our needs may be detrimental not only to us as we are now, impairing our present existence, but also by way of our future development, stunting our growth and development of potentialities, but again, without necessarily positively injuring or harming us. Thus the account implicit in Feinberg's statement should also be modified to take note of that which is detrimental to our *natural development* as individual persons as well as to us as we are now.

Nonetheless, not all that is detrimental to what is commonly called self-development relates to needs, for much so-called self-development is self-activity, self-making. What are properly to be called luxuries in one sense may be necessary for the successful engagement in such self-activity, self-making. Even within what is properly called natural self-development, the lack of that which is detrimental to the individual's development, may not always properly be called a need. For example, a child on a remote, inland farm may potentially be a world-class swimmer, and this fact may be known. The only way of providing for this development, whilst not harmfully affecting other avenues of the child's development, may be to build an olympic size pool for her on the farm, and this neither the farmer nor the state can readily afford to do. I suggest that the child cannot properly be said to have a need for the swimming pool, that it is something desirable, but more akin to a luxury than to a need. This is more evident if we slightly change the example and make it that the child is simply capable of becoming an accomplished swimmer, and better as a swimmer than at any-

thing else she is capable of doing or becoming. The pool is then quite evidently not a need but a luxury. This is because needs relate to valuable existence and to valuable self-development. If lack of the pool led to the child having a nervous breakdown, deep depression, or even insanity, we should still not be entitled to claim that she had a real need for the pool, although she would need something, namely, medical care and treatment. Again, this is because needs relate to valuable existence and development, not to any existence and development, nor to any self-making, self-activity, but to *development*.

At this point it ought to be noted that the concept of needs with which I am concerned here is not that from which most needs talk in social philosophy and the social sciences is derived, that of the biological sciences, where we can talk in a value free way about the needs of the influenza virus, a cancer cell, or of a person's tape worm, but of needs as they arise and are discussed in normative sciences such as sociology, political science, and to a significant degree, psychology, as well as in moral and political philosophy. It is in terms of this concept of needs, that I wish to suggest that talk about human needs, to be meaningful and useful, must be spelt out in terms of natural goods, goods based on human nature and human ends, where the ends are determined by our natures, either in terms of movement from the potential to the actual in the sense evident in much liberal writing, or in the teleological, purposive sense, as in Thomist natural law ethics. Indeed, it is in the context of the latter theory that we can most evidently set out a meaningful, coherent account of what constitutes a need, and what are basic needs. There are, however, considerable difficulties of a metaphysical and ethical nature which impress me as fatal to that whole approach to human needs, which I cannot explore here—I have sought elsewhere, as in my *Meta-Ethics and Normative Ethics* and *God and Evil* to show that difficulties arise which relate to the determination of the end, its nature and even existence, for the latter would seem to be dependent, ultimately, on the existence of God.

If I am right in rejecting the natural law ethic as false, and hence as the proper home of human needs theory, we must look to human nature theories which involve the idea that man is born with a nature which has potentialities for development which are good and such that these potentialities can be realized only if certain conditions are realized, where the potentiality to actuality model

is that of the acorn to the oak. Where the conditions are such that the possessor of the nature with potentialities would be *impaired, marred, stunted*, as a person if the conditions are not met, we speak of his needs and the talk about impaired, marred, stunted, is to be explained in terms of what is natural to his human nature. This contention is qualified, as in the example concerning swimming, by our judgment about the importance or comparative importance of the good to be enjoyed or developed into existence. It would possibly not be an entirely improper extension of talk about needs, to go on to speak of needs as relating to whatever is necessary for the full development of a person as a human being, and as that particular person with the specific nature he has as that unique person.

However, we in fact do not usually speak of needs but of *interests* and *rights* in the latter kind of case, of rights, ideal, abstract, *prima facie*, rather than of immediately demandable, operative rights which can properly and reasonably be demanded here and now. This, I believe, is in part because the more remote aspects of so-called human development and self-perfection impress me as falling less under the characterisation as development of potentialities on the acorn-oak model, where we (as well as our society) determine within limits set by nature, what we become. We call the good developments, self-developments, self-perfections, and the bad, self-stultifying, self-destructive activities. We speak of the rights to the former, and not so readily—as I have argued elsewhere—to the latter; and it is of rights not needs that we speak, because we grasp that after the basic requirements of existence and physical and, to a lesser extent, psychological well-being are met, we are operating with a different concept, that of *self-making*, rather than engaging in the activity of making the potential actual, moving from potentiality to actuality as do the acorn which becomes an oak, the tadpole a frog.

I am therefore suggesting that, *ceteris paribus*, needs are things which ought, where possible, to be available, not withheld, prevented, and indeed, be supplied where necessary; that where needs cannot be met, society or the world ought to be reordered so far as possible, so that they are capable of being met, or obtained by the person with the need, provided that greater goods are not thereby lost or jeopardized; that talk of human needs and needs of particular persons involves reference to natures, the perfection, development, non-impairment of which are good. That is to say, needs talk involves refer-

ence to a natural good existence and development as human beings from potentiality to actuality; and equally of particular persons; where that, the lack of which leads to a person being impaired, stunted, constitutes a need. If there were no such good natures, we could not indicate human needs. Further, if human existence, and the development of the potential into the actual of human and individual natures is not good, talk of needs becomes inapplicable.

To consider the latter point, that needs and good existence and natures are logically, conceptually connected: I have already touched on the point that if a drive related to something bad or worthless, we should not relate it to a need. Consider here the possibility that all human beings have a sadistic drive, such that they lacked inner contentment and fulfilment if they could not bring harm and suffering to others, and such that other aspects of their nature flourished and flowered when they successfully realized their sadistic drive. I suggest that it would be inappropriate to speak of a need to be sadistic, and this, because the notion of a need relates to a good. Another example might be seen to support this. A gravely ill person who is suffering excruciating pain has a need for treatment and medicines to restore his health. We should have no hesitation in speaking of needs in such cases. Yet, if the medicine can simply prolong life pointlessly for two weeks, and this at a cost of intensifying the suffering, and if the dying person has no wish to continue to live, and cannot achieve anything worthwhile by having his life prolonged, we should be wrong to speak of his need for the relevant medicines and treatment. This is because we do not see the prolongation of his life under these conditions as a good. The example of the drug addict further confirms this. It is sometimes said that the addict has a need for his drug, heroin, alcohol, cocaine, whatever it is. This is because, without it, he is seen to be an impaired person, even more impaired than with it—he loses his ability to engage in valued activities, and those who speak of his need here are thinking of the evils temporarily to be eliminated, the good temporarily to be achieved. However, to speak of needs here is usually to speak quite inaccurately. Take the case of the alcoholic who desperately needs his slug of whisky or quart of beer to get started on the day's work, and who, if deprived of this, cannot face up effectively to the world. His need is apparent, not real, what might be called a false need, for if it is satisfied and his real need for treatment not met, he

will be killed by his drug, and harmfully impaired physically and mentally on the way. (If he is a poet or musician who can rise to his heights only with the aid of his drug, alcohol, then it may well be contrary to his needs but in his interests to have the alcohol.) With the hooked, incurable heroin addict who is stuck with his addiction for life, and who can lead a valuable, worthwhile mode of life with the help of limited, controlled amounts of his drug, I see no reason to deny that the drug meets a real although artificially created need, a secondary not a primary need of the addict, resulting from a change in his nature as that particular person. This is because it is essential for a good, I have assumed a natural good, made possible by his nature, and such that a kind of existence or development that was natural, would not be possible without the drug because his nature has itself changed.

To look now more closely at the assumption that I claim to be implicit in talk about human needs and personal needs of human and personal natures with natural actualities and potentialities. In an important and obvious sense it is clearly true that there is a common human nature shared by most, although not by all, human beings, just as most but not all human beings have one head, two ears, two legs, etc. Possession of rationality, sentience, emotional capacity, imagination, and the like, are features of human beings, where certain basic avenues of developments are natural, in the sense of being dictated by the nature itself, developments of potentialities into actualities. To deny that most human beings are rational and such that they have potentialities which, up to a certain level at least, admit of natural development comparable with the actualization of a potentiality as with the acorn becoming an oak, seems to me to fly in the face of the facts. (Consider learning to walk, play, talk, etc.) So too with emotionality, imagination, sentience. Up to a certain degree of actualization at least, we should rightly declare one who lacks this development to be unnatural, defective, stunted, impaired in his development. The other side to this claim is that if there were no human nature with potentialities, no natural ends, good avenues of natural development, and similarly, in respect of individuals, talk of needs would be inappropriate and unnecessary. If we could simply change 'our nature' rather than develop it by realizing or actualizing potentialities, we ought to speak of good and bad possibilities of self-activity, and of what is necessary for us to realize them; these possibilities need not be, and

probably would not be common to all men, and hence, the appropriate way or manner of discussing them would be in terms of desirable goods, that is, which of such may be said to be possibilities the realization of which is in the interests of the person concerned, and what constitute impediments and what aids to the realization of these possibilities. We should not speak of needs but of goods and interests here.

Relativity and objectivity: I have spoken of needs as if they are objective, as if human needs must by definition be the same for all men for all time, that there is a relativity, differences in respect of personal natures due to differences between these natures, but not in respect of human needs. I wish now to clarify this suggestion, for obviously there is relativity in respect of human needs as well as with personal needs.

The natural development of men, like that of oaks, is determined in part by their environment. With men, the development occurs in an environment which is both physical and social. This affects what is natural, and hence what constitutes a need. For example, for the European child, there is a need for reading; for the tribal aborigine child, ability at tracking. Their real and common need is mastery of language. Feinberg mentions a need for fuel to heat one's dwelling, and the like. The need is in fact to have a dwelling which satisfies certain conditions—what is necessary to achieve this will depend on the relevant climatic conditions. Similarly with amounts and kinds of food.

Certain roles and stations are thought to be relevant to needs, and, although the needs to which they give rise are secondary, relative needs, they may be spoken of as real needs. Consider here the needs of mothers in different climatic, health environments.

Needs and interests: In this paper I have operated on the basis of an unexplained distinction between interests and needs, and have suggested that if we do not have a human and personal nature, with inbuilt potentialities of natural development, we should drop talk of needs in favour of talk about possibilities of goods, interests, and the like.

Interests impress me as embracing much more than needs. If we have needs which lack fulfilment, it will commonly be in our interests that they be fulfilled. This is because interests relate to a person's good, where the good is determined not

simply by a man's needs, but also by his goals, aspirations, values, and achievements, and where a man may be ignorant of, mistaken concerning, confused about his interests. It may be in a person's interests temporarily or even permanently to curtail the satisfaction of his needs, as in pursuing a religious vocation, a vocation as an artist, musician, philosopher, explorer, today and certainly before the development of the omni-benevolent, welfare state. It may even be more in the interests of a person that he die for his beliefs than live by recanting them—I believe this to be true of martyrs such as Cranmer and Laud if we take note of their beliefs in an after-life, and possibly even if we do not do so. Thus we may be faced with the choice as to whether to promote a man's interests or his needs, even though usually it is the case that it is in a man's interests that his needs be satisfied. (A case where they might conflict is where artistic inspiration comes only with malnutrition.)

The concepts of *well-being* and *welfare* are obviously related to but distinct from those of needs and interests. They are commonly used interchangeably but also in ways which allow a distinction to be drawn between well-being and welfare. I see well-being as relating to an individual's good, and hence often, to his effective development, where his goals, aspirations, values, and achievements are less decisive in determining what relates to his well-being than to his interests. What is dictated by a person's well-being may conflict with what is dictated by his interests as that particular person—consider the cases cited above. N. Rescher explains welfare in terms of physical welfare (health), material welfare (prosperity), and spiritual or psychological welfare (state of mind or mental health), observing: "Physical health, material circumstances, and mental and emotional well-being are the key elements of welfare." (P. 4)³ And: "Happiness, clearly, is not a component of welfare, but its goal." (P. 5). And: "Welfare is a matter of the resources that a man needs for at least minimal success in 'the pursuit of happiness'." (P. 7). This would appear to be the way in which welfare is conceived of today. Hence it will usually involve satisfaction of needs and more. Typically, promoting a person's welfare will be to act in his interests and to assist in respect of his well-being; but this need not always be so. It may be in a person's interests to forgo happiness and the conditions of happiness. Similarly, well-being may be achieved

without happiness and all the conditions necessary for happiness. The well-being of some persons, for example Socrates, seems to involve a restless, unsatisfied pursuit of goods such as knowledge, without the possibility of happiness, and such that organizing things for their happiness may interfere with their well-being as the kinds of persons they are.

Rights and needs: As I have sought to explain needs, they are obviously relevant to rights. As a general, ideal, *prima facie* right, we have a right to the satisfaction of those needs which human action can cause to be satisfied or thwarted. Thus we have a right to our needs for food, shelter, air, society of other human beings, to sleep, dream, to wakefulness, exercise, rest, and the like, because, and in so far as these relate to goods. We do not have absolute, or, as I should prefer to speak to them, operative rights to the satisfaction of all our needs because, besides this being impossible because of conflict between them due to limitation of resources as well as for other reasons, the *prima facie* right may be overridden by rights to higher, more important goods relating to interests, ideals and values. We do not have a right to the satisfaction of those needs which it is impossible to have satisfied by human action, and where human action does not frustrate the realization of the need, for example, in respect of the need for affection. (In so far as human interference may bear on the realization of the need for affection, it is likely to involve violation of more basic rights such as those of freedom and self-creativity in the sense of freedom to realize possible goods.) Further, we have rights to much more than simply the right to the satisfaction of our needs, even if needs could be extended, as attempts are made to extend them, to include things such as needs for private possessions including land, privacy, respect, education, pleasure and happiness. Human beings have rights which are *pre-need* and *post-need*. We have a right to life but not a need to or for life, for needs here relate to what is essential for valuable life. We have a right to liberty, but the right to liberty goes far beyond what is dictated by needs. Consider here Mill's arguments for liberty in terms of knowledge, rationality, one's interests ('each is the best judge and guardian of his own interests'), and goods attainable by us. The right to liberty relates to possibilities of self-realization, self-creation, even

³ Nicholas Rescher, *Welfare* (Pittsburgh, 1972); see also ch. 9, fn. 2.

more than to needs. This applies also to such rights as those to education, family, property, privacy—the goods involved are not restricted to those involved in needs and need-satisfaction. Further, even where needs are relevant, needs and need-satisfaction would justify too little. This is true in respect of needs for and rights to food, drink, shelter; we clearly have a right to much more than would be dictated by satisfaction of needs.

Needs, rights and roles: I have been concerned with what might be called primary needs and rights. However, much talk of needs, and of rights related to them, concern roles, and needs that are bound up with roles. We speak of the needs of parents, children, and by extension which is not strictly accurate, we sometimes speak of other roles as creating needs, roles such as those of teacher, scholar, judge, carpenter, ruler, farmer, etc. The reference to needs is most at home with roles which are based on human nature, as with parent, mother, etc. With other roles, which relate simply to possibilities and not to potentialities in the sense explained, we should be better advised to speak of goods, conditions essential to the carrying out of duties of the role, and the like, of what is in the interests of the role-occupier as such, rather than of needs, for not all roles, and not all that is necessary for the successful occupancy of a role is a need. It is noteworthy that we are not at all inclined to extend talk about needs to roles which we consider to be unnatural; and so too with bad roles. We do not speak of the needs of a person who is by nature male, who has adopted a female role in life, in respect of that role. Similarly in respect of the 'needs' of the professional thief, murderer, black-mailer, drug pedlar, forger, *qua* their roles, although it is true that we can speak of their interests as such. Rather, it is with what are thought of as natural, good roles, that we most easily speak of needs; this is extended to what is strictly inaccurate talk of needs relating to useful, non-evil roles which are thought to be necessary for the realization of possibilities but not potentialities for good. In the latter cases, it is of goods, interests, and even luxuries necessary for the carrying out of the role, or which make the role or its ends attainable, which ought to be referred to, and not needs.

Needs, rights and political values: Needs as I have sought to explain them are obviously relevant to

political values such as pleasure and the reduction of suffering, happiness and the elimination of misery, in so far as the latter may be distinguished from the former. However, needs are not the only factors which bear on human pleasure and happiness, human pain, suffering and misery, and our concern here is with these goods and evils, and with needs as relevant to them. Self-development and self-perfection are other values which cannot properly be explained without reference to needs, but in so far as they are to be elucidated entirely in terms of satisfaction of needs, they relate only to some of the values with which those who have referred to these as major political values have been concerned. Liberals have been able to neglect to refer to needs to the degree they have, because they have been concerned with something much more comprehensive than merely development of potentialities into actualities. They have been concerned with self-activity leading to the creation of goods, the extension of knowledge, rational belief, culture, artistic creation, and the like. Hence it is that they have been concerned to stress goods and interests rather than needs. Other political values also involve references to needs, for example, justice. Yet when we seek to explain justice by reference to needs, we find that it does not take us very far; what is important as being just and fair, and as involving significant equality, is procedures and distributions which involve considerations that may not always involve reference to needs, and where they do, go far beyond them. Justice in the administration of the law, political justice, racial and religious justice, justice in the distribution of property and rewards for labour, and the like, involve reference to needs as only one factor among many, where they involve it at all; much more they involve concern with goods and values such as human happiness, liberty, respect for persons and desert. The problem of explaining and justifying the emerging value of the twentieth century, privacy, has led some to seek to explain and justify it in terms of there being a need for privacy—consider here Westin's discussion.⁴ Such accounts are unconvincing and unsatisfactory. I suggest that here too, that if privacy is to be shown to be a value at all, it will be in terms of other values such as human happiness, justice, and respect for persons, and not in terms of some previously unrecognized need. Fraternity has not figured very prominently as a political value in any explicit way. I see a

⁴ See A. F. Westin, *Privacy and Freedom* (New York, 1967).

feeling towards acceptance of fraternity as a value among those who deplore competition and seek its replacement by cooperation, who aspire to a society in which envy is absent, and good will towards others prevails. If I am right in this interpretation of fraternity, then satisfaction of needs is relevant to the vindication of fraternity as a political value—if competition is more conducive to the satisfaction of needs than is uncompetitive cooperation, as J. S. Mill inclined to believe in spite of his Taylor-made socialism, then fraternity is less evidently a political value; if the contradictory is true then the case for fraternity is stronger. If envy relates to relative deprivation, to relativities as much as or even more than to need-satisfaction and dissatisfaction, again, to that extent, fraternity and the elimination of envy and promotion of good will, will take us far beyond concern with needs. It may even involve the moral

education of the envious by the state in respect of needs and values and justice in social arrangements, for the envy that causes ill-will and social discontent in welfare states is not obviously related to injustices any more than to need dissatisfaction. Earlier in this paper I indicated how concern for liberty as a political value is and involves a concern for values far more encompassing than simply those involved in needs. I could go on looking more closely at these political values, and their relation to needs, and at other values that have been advanced as genuine political values. I hope that enough has been said here to indicate the general direction of the argument I should seek to develop here, if I were to seek to make good my case that political values and needs, whilst related, are much less tightly connected than is so commonly claimed today, especially by anti-liberal thinkers.

La Trobe University

Received February 25, 1975

II. THE NOTION OF A PRACTICAL INFERENCE

RONALD D. MILO

INTRODUCTION

IN attempting to give an account of what it is for a person to have a reason for doing something, a number of philosophers have introduced the notion of a practical inference. They have supposed that if we can speak of a person as having a reason for doing something in the same sense of "reason for" as that in which we can speak of someone as having a reason for believing something, then, just as there are warranted inferences terminating in beliefs, so must there be warranted inferences terminating in actions—or, at least, resolves (intentions or decisions) to act. Assuming that having a reason for believing some proposition consists in believing some premiss(es) on the basis of which one is warranted in inferring that that proposition is true, they have attempted to explicate the notion of having a reason for performing some action in terms of being warranted in inferring that that action is to be done. The latter kind of inference is said to be practical since, just as my inferring that *P* is true involves my believing *P*, so my inferring that *D* is to be done by me involves my doing *D* or, if now is not the appropriate occasion, my resolving to do *D*. Hence, these philosophers have attempted to show how there can be warranted practical inferences.

According to this view, the generic notion of *having a reason for*—believing *P* or doing *D*—is that of *being warranted in inferring that*—*P* is true or *D* is to be done; and the notion of having a reason for doing something is to be explicated by showing how and in what sense one can be warranted in inferring that something is to be done. I shall argue that this view—I shall refer to it as *the inferentialist position* and to those who subscribe to it as *inferentialists*—is untenable. I shall try to show that when the notion of having a reason for doing something is explicated in this way, that notion loses its normative import for assessing a person's actions as reasonable or unreasonable, because one is attempt-

ing to apply to actions criteria appropriate only to the assessment of the reasonableness of beliefs. I shall first develop the notion of a warranted practical inference and the correlative notion of a valid practical argument (section I). Then I shall examine and criticize several criteria for determining the validity of such arguments that have been proposed by various philosophers (sections II–III). Finally, I shall point out what I believe to be the common error underlying any inferentialist account of having a reason for doing something and shall make some suggestions towards developing an alternative non-inferentialist account of reasons for doing (section IV).

I

In examining the inferentialist position, I shall confine myself to a consideration of those cases where a person may be said to have a reason for doing something in virtue of his wanting something. We commonly believe that our wants provide us with reasons for doing things. If *A* wants to lose weight, *A* has a reason for dieting. If *A* wants to become a physician, *A* has a reason for studying medicine. In such cases *A*, in virtue of his wanting a certain state of affairs *E* to be realized, is thought to have a reason for performing a certain act *D*. According to the inferentialist, if this commonplace conviction is true, then we must be able to discover an inferential relationship between *A*'s wanting *E* to be realized and *A*'s doing or resolving to do *D*, such that anyone who wants *E* to be realized (and, as we shall see, who has a certain relevant belief) is warranted in inferring that *D* is to be done by him (and hence warranted in doing or resolving to do *D*).

Now the way in which we determine whether a theoretical inference is warranted is by constructing a corresponding argument and then examining this argument to see whether or not it is an instance of a correct deductive or inductive argument-form.¹

¹ As I employ the terms "inference" and "argument," an inference is a psychological process consisting—in the case of theoretical inferences—in a person's coming to believe one proposition on the basis of having come to believe some other proposition or propositions. Corresponding to any possible theoretical inference, one can construct an argument, consisting of a set of propositions which may be thought of as the objects or contents of the belief-states constituting the inference.

Thus, if there can be warranted practical inferences, it would seem that they too can be shown to be warranted by constructing corresponding practical arguments that can be seen to be instances of correct argument forms. In the case of practical inferences, however, the construction of a corresponding argument poses a number of difficulties. First of all, the basic elements of the practical inference alleged to explicate the notion of *A*'s having a reason for doing *D* in virtue of his wanting *E* (to be realized) are not beliefs but rather a desire and an action or resolve to act. This poses a problem concerning what the elements of the corresponding argument are supposed to be. We can say that, just as corresponding to beliefs there is what is believed (propositions), so corresponding to desires and actions or resolves to act there is what is wanted and what is to be done. But then the question arises as to how these things can be expressed in language.

A number of suggestions have been made concerning how to express what is wanted. Some have suggested sentences employing evaluative terms—e.g., "It would be good (for me) to lose weight" or "It would be desirable (for me) to lose weight." Sentences in the imperative mood have also been suggested—e.g., "Lose weight" or "Let me lose weight." And some have proposed sentences in the optative mood—e.g., "Would that I lose weight" or "Let it be the case that I lose weight." The rationale behind each of these proposals is to find something—an evaluative proposition, an imperative, an optative proposition—such that accepting it (assenting to it) involves wanting something. (Just as accepting a proposition expressed by an indicative sentence typically involves believing something.)

However, none of these proposals seems entirely satisfactory. Evaluative sentences seem more appropriate for expressing value judgments rather than mere states of wanting—to use Dewey's terms, they seem more appropriate for expressing "appraisals" rather than mere "prizings." Imperatives addressed to others often express what is wanted (for example, "Close the door," uttered by someone who wants the door closed). But the kind of self-addressed imperatives that (we shall see) would be needed as premisses of practical arguments seem more appropriate for expressing resolves or decisions to act rather than wants. And it has been suggested that optative sentences express wishes rather than wants.

But perhaps an optative sentence like "Let it be

the case that I lose weight" comes closest to doing the job of expressing my wanting to lose weight (in the sense of expressing what I want, not the fact that I want it). Furthermore, if this proposal is coupled with a proposal to distinguish between two kinds of propositions—optative propositions expressed by sentences like that above and assertoric propositions expressed by sentences of the form "It is the case that . . .," both the objects of beliefs and the objects of wants can be expressed in the same way—*viz.*, by subordinate clauses having the form of complete indicative sentences.

In looking about for something to represent what is to be done, sentences in the imperative mood seem to be as good a candidate as any. In inferring that something is to be done one either resolves to do something or, if now is the appropriate time for doing it and one is able to do it, one does it. What is needed in language to express the result of a practical inference is a sentence such that to accept or assent to what it expresses is to do or resolve to do something. And imperative sentences seem fairly well suited for this job. Perhaps the most natural way to express my resolve or decision to diet is "I *shall* diet." But for purposes of constructing a practical argument the elements of such an argument can be made more analogous by employing imperative sentences of the form "Let me diet" or (better still) "Let me make it the case that I diet."

Let us then opt for the optative "Let it be the case that I lose weight" to express my wanting to lose weight and the imperative "Let me make it the case that I diet" to express my resolve or decision to diet. According to the inferentialist, my wanting to lose weight provides me with a reason for dieting just in case, given that I want to lose weight, I am warranted in dieting or resolving to diet. And this inference is warranted only if the argument having as its premiss, "Let it be the case that I lose weight," and as its conclusion, "Let me make it the case that I diet," is an instance of a correct argument form. Now some philosophers have suggested that this is an enthymeme, and that it can be seen to be an instance of a valid argument form if we supply the missing premiss—namely, "(It is the case that) one loses weight if and only if one diets."

Thus, the (allegedly) valid argument supposed to show that our practical inference is warranted may be stated as follows:

- (1) Let it be the case that I lose weight.

(It is the case that) one loses weight if and only if one diets.²

Therefore, let me make it the case that I diet.

It is in this way that, according to the inferentialist, we are to explicate the notion of *A*'s having a reason for doing some act *D* in virtue of his wanting some state of affairs *E* to be realized and his believing that one can make it the case that *E* if and only if one does *D*.

Let us grant that if (1) is valid, then it is possible to make warranted practical inferences of the sort we have been considering. The problem which now faces the inferentialist is to explain in what sense (1) can be said to be a valid argument. An argument, as ordinarily conceived of, consists of (assertoric) propositions—things that typically can be true or false. And a valid argument is commonly defined as one whose conclusion must be true if its premisses are true. How then can the notion of an argument (let alone that of a valid argument) be applied to so-called "practical arguments" like (1), which consists of things like optative and imperative "propositions" that can be neither true nor false—without, that is, either distorting the meaning of the term "argument" or else using it in a different sense? In order to cope with this problem some inferentialists have suggested that there may be some value other than truth which can be possessed by optative and imperative propositions, and that this value may serve as a substitute for truth. Just as the rules determining the validity of theoretical arguments are designed to ensure that we never pass from true premisses to a false conclusion, so the rules determining the validity of practical arguments are designed to ensure that we never pass from premisses possessing the substituted value to a conclusion that lacks it.

II

Only two kinds of truth surrogates have been seriously considered. One of these is satisfaction (or fulfillment).³ An optative or imperative proposition may be said to be satisfied when the state of

affairs described by its subordinate clause is realized. For example, the optative proposition, "Let it be the case that I lose weight," is satisfied just in case I do lose weight. And the imperative proposition, "Let me make it the case that I diet," is satisfied just in case I diet.

We can now say that a practical argument like (1) is valid just in case its conclusion must be satisfied if both its premisses are satisfied. Is argument (1) valid according to this criterion of validity? Now it is certainly true that if one loses weight if and only if one diets (i.e., if the second premiss is satisfied) and if I do lose weight (i.e., if the first premiss is satisfied), then I must have dieted (i.e., the conclusion must be satisfied). So argument (1) does appear to be valid according to this criterion.

Unfortunately, this criterion allows us to show how (i.e., in what sense) arguments like (1) can be valid only at the cost of making them no longer practical. If the validity of a practical argument consists in the impossibility of its premisses being satisfied unless its conclusion is also, then, if the premisses are satisfied, one who believes that they are satisfied cannot go wrong in believing (on this basis) that its conclusion is satisfied. But this is just to say that he cannot go wrong if he makes a certain *theoretical* inference. Now the "inference" which the validity of (1) is supposed to warrant consists in a person's coming to resolve to diet (or actually engaging in dieting) on the basis of his wanting to lose weight and believing that one loses weight if and only if one diets. This "inference" is genuinely practical. But it is not at all evident how the fact that the conclusion of (1) must be satisfied if its premisses are warrants such an "inference." If the premisses of a valid theoretical argument are satisfied (true), one who believes them cannot go wrong in believing (on this basis) its conclusion. If, on the other hand, the premisses of (1) are satisfied, it is not that I cannot go wrong in dieting or resolving to diet—rather, I have already dieted.

Moreover, even if one grants that there can be warranted practical inferences, this view surely misconstrues what must be the point of making

² This premiss is stronger than would be required by many models of a valid practical argument. Some require only a premiss stating a necessary condition (one loses weight only if one diets), while others require only a sufficient condition (if one diets one loses weight). I choose the stronger premiss for my example, since it gives me an argument that is valid according to all of the inferentialist views I shall be discussing.

³ This sort of view was proposed and elaborated by Albert Hofstadter and J. C. C. McKinsey in their "The Logic of Imperatives" (*Philosophy of Science*, vol. 6 [1939], pp. 446ff.). It is now generally regarded as unacceptable because, as Alf Ross pointed out in his "Imperatives and Logic" (*Philosophy of Science*, vol. 11 [1944], pp. 30ff.), the criterion of validity utilizing this value has counter-intuitive or paradoxical consequences. However, since the real defect of this view—which is the source of the counter-intuitive consequences—has not been fully appreciated (even though Ross, I think, was aware of it) and since a similar defect attaches to the other criteria of validity that I shall discuss, I think it is still worth discussing.

such inferences. Although at least one of our aims in making theoretical inferences is to pass from true beliefs (i.e., beliefs whose propositional objects are satisfied) to other beliefs that are also true, it could not be one of our aims in making practical inferences to pass from wants whose propositional objects are already satisfied (i.e., wants for ends which we have already attained) to assents to imperatives whose propositional objects are satisfied (i.e., to performing acts or resolving to perform acts that we have already done).

Because of this some have suggested that we choose another surrogate for truth—one more appropriate to the point of making practical inferences. Thus it has been suggested that the value which we aim to preserve in making practical inferences is that of satisfactoriness (or appropriateness).⁴ Unlike truth, the value of satisfactoriness is a relative notion—and doubly so. An optative or imperative proposition is always satisfactory *for* some end or set of ends and *to* some person (persons) who desires (desire) the end or set of ends.

Satisfactoriness may be defined in terms of the notion of satisfaction: an optative or imperative proposition is said to be satisfactory relative to a given end or set of ends just in case that end or set of ends is realized if that proposition is satisfied. An optative proposition that expresses a desire that a certain end be realized will be trivially satisfactory relative to that end, since obviously that end is realized if the proposition is satisfied.

Now the criterion of validity suggested by this proposal to substitute satisfactoriness for truth seems to be as follows: a practical argument is valid if and only if its conclusion must be satisfactory if its premisses are. This, however, cannot be applied, as it stands, to an argument like (1), since it contains an assertoric proposition which cannot have this value. Presumably, a practical argument like (1) will be valid just in case its imperative conclusion must be satisfactory if its optative premiss is satisfactory and its assertoric premiss is true. Generalizing, we may say that a practical argument is valid if and only if its conclusion must be satisfactory if its non-assertoric premiss(es) is (are) and its assertoric premiss(es)—if it has any—is (are) true.

If the satisfaction criterion of validity is unacceptable because of its paradoxical consequences, then it must be noted that the satisfactoriness criterion yields paradoxical consequences of its own.⁵ Perhaps, however, the most serious objection to adopting the satisfactoriness criterion of validity is that the validity, *in this sense*, of argument (1) does not guarantee that one who accepts its premisses has a reason for accepting its conclusion. Argument (1) is valid by this criterion just in case its conclusion must be satisfactory if its premisses are, respectively, satisfactory and true. Now in order for *A* to have a reason for dieting in virtue of his accepting the premisses of (1) is it sufficient that the conclusion of (1) must be satisfactory *if* its premisses are, respectively, satisfactory and true, or is it necessary that its conclusion actually be satisfactory?

It seems obvious that *A* has a reason for dieting only if the conclusion of (1) is actually satisfactory for *A* relative to the end of losing weight. And this will be so only if the assertoric premiss of (1) is not merely believed by *A* to be true but is actually true. For, if one's dieting is not a sufficient condition of one's losing weight (even though *A* may believe that it is) how can *A*, in wanting to lose weight, be said to have a reason for dieting? All that follows from the validity (in this sense) of (1) is that if *A* wants to lose weight and *believes* that the assertoric premiss is true, he is thereby warranted in *believing* that the conclusion of (1) is satisfactory—and hence in *believing* that he has a reason for dieting. Here again, only a theoretical inference is warranted. Thus, *A*'s having a reason for dieting in virtue of his accepting the premisses of (1) requires not only that (1) be valid but also that its assertoric premiss be true. Moreover, as I shall later argue, *A*'s having a reason for dieting in virtue of his wanting to lose weight requires only that the assertoric premiss of (1) be true—it is not necessary that *A* believes (accepts) it.

III

Instead of attempting to define validity in terms of some surrogate for truth, such as satisfaction or

⁴ This proposal is made by A. J. Kenny in "Practical Inference," *Analysis*, vol. 26 (1965-6), pp. 65ff. A rather similar proposal is made by H. N. Castaneda, "Imperative Reasonings," *Philosophy and Phenomenological Research*, vol. 21 (1960), see pp. 35ff. Although I shall be discussing the view proposed by Kenny, my basic criticisms of Kenny's proposal apply also to Castaneda's proposal.

⁵ Kenny (*op. cit.*) himself admits that this criterion warrants certain inferences "which *prima facie* appear paradoxical"—perhaps the most notorious of these being the inference from "Open the door" to "Open the door and smash the window." Kenny tries to explain away their paradoxical appearance (see p. 74). But see R. M. Hare, *Practical Inferences* (London, 1971), pp. 65-66, for a criticism.

satisfactoriness, other inferentialists have proposed a kind of "psychological" criterion of validity.

Sometimes we say that an argument is deductively valid if and only if if the premises are true then the conclusion must also be true. But sometimes we say instead that an argument is deductively valid if and only if after all the deductive steps have been set out for a man he doesn't believe the conclusion to be true, then he can't really believe the premises to be true. He can say that he does—but he can't really be doing so. Sometimes we use this latter account of deductive validity to explain what the former account *means*; sometimes we say that the latter account is justified by the former account. In any case there are these two possible accounts—and if there is no parallel to the former in the case of practical reasoning, there is at least something very like the latter.⁶

According to this criterion, the validity of an argument consists in the logical impossibility of one's accepting (assenting to) its premiss(es) but *not accepting* its conclusion. An alternative version of this criterion identifies the validity of an argument with the logical impossibility of one's accepting its premiss(es) but *rejecting* (dissenting from) its conclusion.⁷

The latter version of this criterion seems to be what Hare had in mind when, in *The Language of Morals*, he defined entailment in terms of the notions of assent and dissent. According to this criterion an argument (whether theoretical or practical) is valid (i.e., one whose premiss(es) entail its conclusion) if and only if the fact that a person assents to its premiss(es) but dissents from its conclusion is a sufficient criterion for saying that he has misunderstood at least one of them.⁸ Thus, a theoretical argument is valid just in case it is logically impossible for one to believe its premiss(es) while disbelieving its conclusion—unless one has misunderstood at least one of the sentences expressing them. But, of course, if one has misunderstood the meaning of at least one of the sentences expressing the premisses, then, whatever one believes when one supposes oneself to be believing them, one does not *really* believe them. And the same may be said of the conclusion. Thus, this criterion may be expressed more simply in

terms of the logical impossibility of believing the premisses while disbelieving the conclusion.

Notice that, according to this criterion, it is not merely logically *impermissible* for one to disbelieve the conclusion of a valid argument while believing its premisses, it is logically *impossible*. In a later essay Hare explicitly adopts as his criterion the *logical impermissibility* of assenting to the premiss(es) but dissenting from the conclusion. Granted that a reasoner intends to avoid contradicting himself, the rules of (deductive) logic tell him how to do this.

If I assent to '*p* and *q*' (where '*p*' and '*q*' are assertions), I am bound not to dissent from '*p*', *if and only if I am to avoid inconsistency*. Of course if (like Walt Whitman) I do not care about inconsistency, I can say '*p* and *q* and not *p*' with equanimity. In the same way, I am bound, having said 'Do *a* and do *b*', not to say the contradictory of 'Do *a*', if and only if I am to avoid inconsistency.⁹

Hare admits here that, although it is logically impermissible for one who assents to "Do *a* and *b*" to dissent from "Do *b*," it is nevertheless logically possible. But he fails to remember that his criterion for assenting to commands makes this, not just logically impermissible, but logically impossible.¹⁰ And the same is true of assenting to "*p* and *q*" while dissenting from '*p*'. For, according to Hare's criterion for assenting to assertions or statements, this involves not just *saying* "*p* and *q* and not *p*," as Hare suggests above, but *believing* what "*p* and *q*" expresses while *disbelieving* what '*p*' expresses. If one *said* this, that would be a sufficient criterion for saying that one did not understand the meaning of "and." In that case one could not really mean (and thus could not really believe) what one said. Thus, it is logically impossible to believe "*p* and *q*" while disbelieving '*p*'. Moreover, it is logically impossible to *disbelieve* '*p*', given that one believes "*p* and *q*" just because it is logically impossible *not to believe* '*p*'.¹¹ Thus, the second version of the proposed "psychological" criterion of validity collapses into the first version—namely, that an argument is valid if and only if it is logically impossible to accept (assent to) its premiss(es) but not to accept its conclusion.

⁶ Judith Jarvis Thomson, "Practical Reasoning," *The Philosophical Quarterly*, vol. 12 (1962), p. 317. She does not herself, however, build a theory of practical inference on this observation since she denies that there is any such thing.

⁷ Thus, we do not have here a true psychological criterion of validity. The criterion is not the psychological impossibility of certain states of affairs obtaining, but the logical impossibility of certain psychological states of affairs obtaining.

⁸ R. M. Hare, *The Language of Morals* (Oxford, 1952), pp. 25, 172.

⁹ *Practical Inference* (op. cit.), p. 73.

¹⁰ Cf. *The Language of Morals* (op. cit.), p. 20.

¹¹ Because: "*A* believes *P* and *Q*" entails "*A* believes *P* and *A* believes *Q*."

When this "psychological" criterion is applied to a practical argument like (1), such an argument may be said to be valid just in case it is logically impossible for a person to have the want and the belief involved in assenting to the premisses while failing to assent to the conclusion, i.e., failing to perform or resolve to perform the act mentioned in the conclusion.¹² Unfortunately, argument (1) does not seem to be valid according to this criterion, since the psychological state of affairs just described does not seem to be logically impossible. It is more plausible to suggest that the following is logically impossible: *A* wants to attain *E* more than he wants to attain anything else, and *A* believes that he (one) will attain *E* if and only if he (one) does *D*, yet *A* neither does nor resolves to do *D*. Supposing this to be true, then argument (1) would be valid by this criterion if we construe its first premise in such a way that assenting to it involves wanting to lose weight more than one wants anything else. The trouble with this is that very few practical inferences would be warranted by such an argument, since very few persons could assent to such first premisses.

A more serious difficulty with this proposed "psychological" criterion of validity, even when applied to theoretical arguments, is that it robs the notion of having a reason for believing or doing something of its normative import. And like the satisfaction and the satisfactoriness criteria, it fails to make the validity of a practical argument relevant to the reasonableness of accepting its conclusion, given that one accepts its premiss(es).

In the case of some very simple arguments, i.e., those whose validity is established by a single simple rule of inference, it does seem impossible for a person to believe the premiss(es) while not believing the conclusion. For example, this seems to be true of arguments having the form "*p* and *q*, therefore *p*." And it has been suggested that it is also true of arguments valid by *modus ponens*.

When a man thinks that *P* and also that if *P* then *Q*, it is logically impossible for him not to think also that *Q*: he may of course assert the premises explicitly without uttering the conclusion, but it is impossible

that he should fail to think *Q*—say it 'in his heart' as it were.¹³

Now whatever else one may think about the merits of taking the logical impossibility of such a psychological state of affairs as a criterion of the validity of *modus ponens* arguments, it is obvious that adopting such a criterion has the consequence of robbing rules of inference of their normative force by making it logically impossible (not merely logically impermissible) for anyone to contravene them.¹⁴ And insofar as the inferentialist makes *A*'s having a reason for believing *P* in virtue of his believing *Q* depend on the fact that "*Q*, therefore *P*" is a valid argument, then (at least in the case of simple valid arguments, i.e., those valid by a single simple rule of inference, such as *modus ponens*) it has the further consequence of making it logically impossible for one to have a reason for believing something without *already* believing that which he has a reason to believe.

Unless it is logically possible for one to either fail to believe or to disbelieve the conclusion of a valid argument, given that one believes its premiss(es), there will be no point in speaking of the unreasonableness of disbelieving the conclusion. The fact that it is logically impossible for the premiss(es) of an argument to be true but its conclusion false does have a normative bearing on the reasonableness of believing the conclusion, given that one believes the premisses. But this will be so only insofar as it is possible for one not to believe the conclusion—and thus to believe its contradictory. However, when the validity of an argument is identified with the logical impossibility of one's not believing its conclusion, given that one believes its premisses, the validity of an argument fails to have this normative import.

IV

I should like to conclude by pointing out what I believe to be the fundamental error underlying any inferentialist account of one's having a reason for doing something in virtue of one's wanting some end to be realized. All such accounts confuse what

¹² G. H. von Wright appears to adopt this sort of criterion in "Practical Inference," *The Philosophical Review*, vol. 72 (1963), pp. 165–166.

¹³ Max Black, "The Gap Between 'Is' and 'Should'," *The Philosophical Review*, vol. 73 (1964), p. 177.

¹⁴ Actually, this view treats rules of inference as linguistic conventions governing the use of terms like "and" and "if . . . then." It is possible to contravene *modus ponens* by saying "if *p* then *q* and *p*, but not-*q*" but it is not possible to contravene it by *believing* this. Thus the ultimate criteria of validity are linguistic conventions. The logical impossibility of certain psychological states of affairs—e.g., believing, at one and the same time, both *p* and *q* and not-*p*—is a consequence of the fact that if one said "*p* and *q* but not-*p*" one could not mean what one said.

may be called a *reason-making condition* with a constituent of *that in virtue of which one has a reason*.

In order to make clear the nature of this confusion, let us see what such a confusion would amount to in the case of reasons for believing. If *Q* entails *P*, then if *A* believes *Q*, *A* has a reason for believing *P*. Here, the reason-making condition is the fact that *Q* entails *P*, i.e., it is the fact that *Q* entails *P* which makes *A*, in believing *Q*, have a reason for believing *P*. That in virtue of which *A* has a reason for believing *P* is *A*'s believing *Q*. Now if someone held that *A*'s having a reason for believing *P* consists not only in *A*'s believing *P* but also in *A*'s *believing* that *Q* entails *P*, such a person would be treating the reason-making condition of *A*'s having a reason for believing *P* as if it were a constituent of that in virtue of which *A* has a reason.

That this is an error can be shown as follows. Suppose that *Q* is a proposition of the form "All *X* are *Y* and all *Y* are *Z*" and *P* is a proposition of the form "All *X* are *Z*." In that case *A*, just in believing *Q*, has a reason for believing *P*. It is not a necessary condition of *A*'s having this reason that he *believes* that *Q* entails *P*; it is sufficient that *Q* does entail *P*.

Now suppose that *A* wants some state of affairs *E* to be realized and that *E* can be realized if and only if a certain act *D* is done (let us abbreviate this as "*E* requires *D*"). In this case, does *A*, just in wanting *E*, have a reason for doing *D*? It seems to me that the correct answer is "Yes," that *A* has a reason for doing *D* given that *A* wants *E*, just because *E* requires *D*. Here again, I suggest that it is not a necessary condition of *A*'s having this reason that *A* *believes* that *E* requires *D*; it is sufficient that *E* does require *D*.

But if so, then the inferentialist's explication of *A*'s having a reason for doing *D* in terms of *A*'s being warranted in inferring that *D* is to be done by him (i.e., being warranted in assenting to the imperative, "Let me make it the case that I do *D*") on the basis of his wanting *E* (assenting to "Let it be the case that *E*") and *believing* that *E* requires *D* (assenting to "It is the case that *E* requires *D*") cannot be correct. If *A*'s believing that *E* requires *D* is not a necessary condition of *A*'s having a reason for doing *D*—given that he wants *E*, but if nevertheless that *E* requires *D* is a sufficient condition of *A*'s having a reason to do *D*—given that he wants *E*, then the fact that *E* requires *D*, like the fact that *Q* entails *P*, is a reason-making condition. But whereas, corresponding to the fact that *Q* entails *P*, there is a valid argument—"Q,

therefore *P*" (and hence a possible warranted inference, consisting in *A*'s coming to believe *P* on the basis of his believing *Q*) there is no valid argument (and hence no possible warranted inference either) corresponding to the fact that *E* requires *D*. Thus, whereas reasons for believing are, reasons for doing are not inferential in character.

Of course, the inferentialist will not admit the basic premise on which my argument against him rests. While admitting that *Q*'s entailing *P* is a sufficient condition of *A*'s having a reason for believing *P* in virtue of his believing *Q*, he refuses to admit that *E*'s requiring *D* is a sufficient condition of *A*'s having a reason for doing *D* in virtue of his wanting *E*. And he refuses to admit this because he refuses to admit that there can be any other reason-making condition except that constituted by some inference-warranting condition, such as that of entailment.

And I suspect that the inferentialist's reason for making the latter assumption is as follows. If, as I suggest, *A*'s having a reason for believing *P* in virtue of his believing *Q* can be explicated in terms of *A*'s being warranted in believing that *P* is true on the basis of his believing *Q*, but *A*'s having a reason for doing *D* in virtue of his wanting *E* cannot be explicated in terms of *A*'s being warranted in inferring that *D* is to be done on the basis of his wanting *E* (and believing that one attains *E* if and only if one does *D*), then we must conclude that we cannot be said to have reasons for doing things in the same sense of "reason for" as that in which we can be said to have reasons for believing things. We must conclude, therefore, that there are not two kinds of *reasons for*—reasons for believing and reasons for doing. Rather there are simply two senses of "reason for." Thus, a univocal sense of "reason for" demands an inferentialist account of reasons for doing.

In response to this I shall suggest that if we can be said to have justifications for doing things, in addition to there being explanations (reasons why) for our doing things, then we can be said to have reasons for doing things in the same sense of "reason for" as that in which we can be said to have reasons for believing. As an alternative to the inferentialist position, I suggest that the generic notion of a *reason for* is that of a justifying consideration in favor of the reasonableness of that for which one is said to have a reason—which may be either a belief or an action. What needs to be shown now is that there is a sense of the terms "reasonable" and "justified" that may be univocally applied to both

beliefs and actions and which can be explicated—at least in the case of actions—without employing the notion of a warranted inference.

Both beliefs and actions may be appraised as reasonable or unreasonable, justified or unjustified. In the case of beliefs, it is essential to our understanding of the notion of a belief that beliefs may be either true or false; and it is essential to our understanding of what it is to appraise beliefs rationally, i.e., as reasonable or unreasonable, that it is desirable to have beliefs that are true and undesirable to have beliefs that are false.

Now if it is desirable that when *A* believes *Q*, *Q* be true, it is also desirable that when *A* believes *P* on the basis of his believing *Q*, his belief that *P* be truth-preserving with respect to his belief that *Q*. The principles of deductive and inductive inference specify how the propositional objects of one's beliefs must be related in order for one belief to be, respectively, necessarily or probably truth-preserving relative to some other belief or beliefs. Thus, in determining whether it would be reasonable for *A* to believe *P* given that he believes *Q*, the fact that *Q* entails or inductively confirms *P* is a relevant consideration, i.e., a reason-making condition.

But the fact that *Q* entails or inductively confirms *P* is a reason-making consideration only insofar as it implies that *A*'s believing *P*, given that he believes *Q*, is a necessary and/or sufficient condition of *A*'s realizing the end of truth-preservation. If this is so, then the ultimate criterion of justification to which we must appeal in appraising beliefs is that of *vindication*, i.e., suitability or appropriateness to achieving certain ends or goals.¹⁵ For if we regard beliefs as reasonable simply because they are deductively or inductively warranted (i.e., simply because they are *validated*) we still face the problem of showing why such "reasonable" beliefs are desirable. But if we regard the truth-preserving character of the principles of deductive and inductive inference as the basis for regarding such beliefs as reasonable, this problem is avoided. Thus, it is the truth-preserving character of our principles of inference that ultimately accounts for the reasonableness of inferentially warranted beliefs.

Now this criterion of justification (i.e., vindication) is one that is applicable to actions as well as beliefs. Moreover, while it is only indirectly applica-

ble to inferentially warranted beliefs (via its direct application to justifying our adoption of the principles of deductive and inductive inference), it is directly applicable to actions. Actions may be directly vindicated by showing their performance to be a necessary and/or sufficient condition of realizing most effectively a certain end or ends. This explains why the notion of an inference is not required in order to explicate the notion of having a reason for doing something. The fact that the performance of some act *D* is a necessary and/or sufficient condition of the realization of some end *E* is necessarily and directly a reason-making consideration for an agent who desires *E*, if we adopt vindication as our criterion or justification; just as the fact that the truth of *Q* entails or inductively confirms that truth of *P* is a reason-making consideration for one who believes *Q*, because this latter fact implies that *A*'s believing *P*, given that he believes *Q*, is a necessary and/or sufficient condition of *A*'s realizing the end of truth-preservation. Thus, in both cases, the reasonableness of *A*'s *X*-ing (believing *P* or doing *D*) is a function of the utility of *A*'s *X*-ing with respect to the realization of certain ends.

One might object that this analysis overlooks an important disanalogy between beliefs and actions. Unlike actions, beliefs do not seem to be the sorts of things that can be intentional, i.e., done for the sake of an end. But it does not follow from this that beliefs cannot be appraised in terms of their utility for enabling us to achieve certain ends. Truth-preservation is a goal whose realization may be promoted by *A*'s believing *P* even though *A* does not believe *P* in order to preserve the truth of his belief that *Q*. Thus, my analysis does not require us to make a false assimilation of beliefs to intentional actions.

One might also object that, whereas the reasonableness of *A*'s doing *D* is a function of the utility of his action with respect to realizing certain ends that *A* desires, the reasonableness of *A*'s believing *P* cannot be said to be a function of the utility of his belief with respect to realizing certain ends that *A* just happens to desire. If, by believing *P*, *A* will preserve the truth of his belief that *Q*, *A* has a reason for believing *P* whether or not *A* desires to preserve the truth of his belief that *Q*. To suppose otherwise would be to impeach the objective character of reasons for believing.

¹⁵ For an elaboration of the notion of vindication and of the distinction between vindication and validation see Herbert Feigl, "De Principiis Non Disputandum . . . ?" in *Philosophical Analysis*, ed. by Max Black (Ithaca, N.Y., 1950).

There are two possible replies to this objection. One might simply decide to brazen it out and contend that if *A* really does not want to be consistent or to preserve the truth of his beliefs, then the fact that *Q* entails *P* and *A* believes *Q* does not imply that *A* has a reason for believing *P*. But if one is unable to swallow this bitter pill, one may simply qualify one's appraisal of the reasonableness of *A*'s belief that *P*. This could be done, for example, by introducing the notions of *epistemic* reasonableness or justification and of having an *epistemic* reason for believing. These notions could then be explicated by stipulating that to appraise *A*'s beliefs as *epistemically* reasonable or in terms of whether *A* has an *epistemic* reason for believing *P* is (necessarily) to appraise his beliefs from the point of view of their utility for achieving certain *epistemic* ends, such as truth and truth-preservingness—irrespective of whether *A* himself desires to

realize these ends. In this way one could make objective or "categorical" appraisals about what *A* ought (has reason) to believe.

One could similarly qualify one's appraisal of the reasonableness of *A*'s doing *D* by introducing, say, the notions of *moral* reasonableness or justification and of having a *moral* reason for doing something. Here again one could explicate such notions by stipulating that to appraise *A*'s action in terms of whether *A* has a *moral* reason for doing it is to appraise it from the point of view of its utility with respect to such ends as the promotion of the general welfare and/or some ideal of justice. This would allow for equally objective or "categorical" appraisals about what *A* ought (has reason) to do. In any case, it is clear that the vindication criterion of justification may be employed to make either subjective ("hypothetical") or objective ("categorical") appraisals of either beliefs or actions.¹⁶

The University of Arizona

Received February 27, 1975

¹⁶ I am grateful to Merrill Salmon, Henry Byerly, and Keith Lehrer for their comments on earlier versions of this paper.

335087

III. NATURE, PLENITUDE AND SUFFICIENT REASON

R. H. KANE

I. INTRODUCTION

IN this paper, I will defend two claims about a well-known principle in Western thought which Lovejoy has called the principle of plenitude, "no genuine potentiality of being can remain unfulfilled."¹ First, I will argue that the principle of plenitude is a logical consequence of the principle of sufficient reason (everything has a reason for its existence) and also a logical consequence of the principle of causality when the latter is conceived as a particular version of the principle of sufficient reason. Second, I will argue that a version of the principle of plenitude plays a heuristic role in contemporary physics analogous to the role played by the causal principle in prequantum physics.

Both these claims are controversial in themselves and their compatibility can be questioned. Consider the first claim. The principle that no genuine potentiality can remain unfulfilled, or equivalently, that every genuine possibility will be actualized or realized, is surely a puzzling principle. There is every reason to wonder why it has been regarded as necessarily true by so many Western thinkers. We cannot resist asking why every genuine potentiality *must* come to be realized, why some possibilities *cannot* remain forever unfulfilled. Since its denial is not evidently self-contradictory, the principle would not appear to be a *logically* necessary truth, if it is necessary in any sense.

Perhaps the principle of plenitude can be defended as a presupposition of all rational thinking or inquiry as the principles of sufficient reason and causality have sometimes been defended. But plenitude seems an even less promising candidate for a presupposition of all rational thinking than sufficient reason or causality. Many philosophers like Kant, who regarded the truth of the principle of causality as a precondition for any knowledge of nature, were not ready to accord the same status

to the principle of plenitude. They would have been surprised to learn that belief in the principle of causality commits one also to belief in a version of the principle of plenitude. But this is what the first claim says.

A derivation of plenitude from sufficient reason (and causality), supporting this first claim, will be presented in section 4 of the paper. Several preliminary remarks are in order about this derivation. First, it will not appeal to theological notions or arguments as in the case of most historical defenses of plenitude. In most traditional accounts, the necessary "fullness" of being is derived from the omnipotence or perfection of God or the One, whose existence is independently proved (Cf. Plotinus, *Enneads*, V, 4). Nor will the derivation appeal to notions of goodness, compossibility and the best world, such as Leibniz employed to establish his limited version of the principle of plenitude. Such notions may enter into particular interpretations of the principles of sufficient reason or plenitude but they are not needed to demonstrate the logical relation between the two principles.

The second claim, that plenitude plays a heuristic role in contemporary physics, is perhaps even more controversial than the first. It is fair to say that most modern thinkers agree with Lovejoy that the history of attempts to apply the principle of plenitude to nature is "the history of a failure." (P. 329.) Although philosophers continue to be interested in the principle, it is for the light it may throw on the views of historically important figures or, in the case of Leibniz, on issues about the logic of possible worlds. A case in point is a thought-provoking recent essay on Leibniz and plenitude by Jaakko Hintikka.² While Hintikka takes issue with Lovejoy on a number of historical points, he agrees that the history of attempts to apply

¹ A. O. Lovejoy, *The Great Chain of Being* (New York, 1960), p. 52. Hereafter references to this work will appear in the text of the paper.

² "Leibniz on Plenitude, Relations and the 'Reign of Law'" in Harry G. Frankfurt (ed.) *Leibniz: A Collection of Critical Essays* (Garden City, 1972), pp. 155-190.

plenitude to nature is the history of a failure and adds a series of persuasive reasons for rejecting plenitude as a principle of nature.

But, as Hintikka says, everything depends on how the principle of plenitude is interpreted and this is where the derivation of the principle from sufficient reason comes in. The derivation to be offered in this paper will suggest an interpretation of plenitude which escapes the usual criticisms and is both historically significant and philosophically interesting.

II. PLENITUDE AND TIME

A distinction concerning plenitude and time must be introduced at the start because of its importance for understanding the derivation. It is characteristic of natural entities from quasars and planets to primates and condors, that they may come into existence and go out of existence at different times and that they may exist at some times but not at others. Does the principle that every genuine possibility must be fulfilled or realized imply that such natural entities or kinds of entities,³ if they are genuine possibilities, must exist at *all* times (nature is "constantly full") or merely at *some* time, past, present or future (all possibilities will be realized "in the fullness of time")? This question suggests the need for a distinction between two different temporal versions of the principle of plenitude. I will call the first of these two versions of the principle (that all genuine possibilities must be realized at all times), the "all times" or "nature is constantly full" version of plenitude and the second (that all genuine possibilities must be realized at some time) the "some times" or "fullness of time" version.

There is, of course, a history behind these two alternatives. Until evidence began to accumulate against the doctrine of the fixity of species in the 18th century and thereafter, the common view regarding plenitude in the biological sphere was that all possible biological species must exist at all times (nature is constantly full). But when evidence in the fossil remains began to suggest that some species had gone out of existence and breeders in

the botanical gardens of Europe were bringing new species into existence, those who wished to save the principle of plenitude turned to the "some times" or "fullness of time" version. They argued that every possible form of organic matter must exist at some time, past, present or future, not necessarily at all times.

This distinction is important because, as it turns out, the "some times" or "fullness of time" version of plenitude is the only temporal version of the principle which is a logical consequence of the principle of sufficient reason. The strong "nature is constantly full" version, which fell out of favor in the 18th century, cannot be derived from the principle of sufficient reason.

III. SUFFICIENT REASON

To set the proof of plenitude in motion, we need first to define the principle of sufficient reason and lay out some of the logical presuppositions of the proof. Spinoza's statement of the principle of sufficient reason is as clear as any. He says: "Of everything whatsoever, a cause or reason must be assigned alike for its existence or its non-existence." (Ethics, I, XI). We will depart from the letter of this statement, though not I hope from its spirit, by defining the principle in the least controversial logical context, viz. in the symbolism of standard first order predicate logic. To do this I suggest that we use the resources of first order predicate logic with two additions, (i) special temporal variables, '*t*', '*t*', etc., whose domain is restricted to moments or intervals of time and (ii) an operator "*S*()" to represent the idea of a sufficient reason. The temporal variables '*t*', '*t*' etc. will be used as subscripts of the predicate expressions, so that, for example, '*P_tx*' may be read "*x* is (or has the property) *P* at time *t*" and "*P_tⁿx₁ ... x_n*" may be read "*x₁ ... x_n* stand in the *n*-adic relation *P* at time *t*."⁴

The operator "*S*()" will be read "that () has a sufficient reason for being the case" and may be applied to any meaningful, or well-informed, formula. Thus "*S*(*P_tx*)" may be read "that (*x* is *P* at time *t*) has a sufficient reason for being the case"

³ When applied to nature, the principle of plenitude has sometimes been applied to particular entities, sometimes to natural kinds. The derivation of section 4 is independent of this distinction since the principle of plenitude to be derived there can be applied either to particular entities or kinds depending on the interpretation preferred. The distinction between two temporal versions of plenitude is another matter, however. For, as we will see, only one of the temporal versions of plenitude can be derived.

⁴ Temporal variables are meant to stand for "stable" time specifications like "5 January, 400 AD" or "3:00 AM, 20 June, 941" and not for "unstable" time specifications like "now," "then," "five days ago" etc.

and " $S(-P_t x)$," "that (x is not P at t) has a sufficient reason for being the case." In this way we can speak of there being sufficient reasons for the existence or non-existence of states of affairs as Spinoza's statement of the principle of sufficient reason requires. Of course, we are limited to the existence and non-existence of *states of affairs*, but this can be remedied in part once quantifiers are introduced. For example, the expression " $S((\exists x)(P_t x))$," (read "that [there exists an x which is P at time t] has a sufficient reason for being the case") asserts that there is a sufficient reason for the existence of some object having the property P at time t .

Such expressions with quantifiers also allow talk of sufficient reasons for the existence or non-existence of *kinds* of things at different times. For example, assuming the predicate ' P ' defined a kind of thing, say, a quasar, primate or reptile, then " $S((\exists x)(P_t x))$ " would assert that (there exist quasars [primates or reptiles] at time t) has a sufficient reason for being the case and " $S(-(\exists x)(P_t x))$ " that (no quasars (primates or reptiles) exist at time t) has a sufficient reason for being the case. Finally, with quantification over temporal variables, one can talk about reasons for the existence of entities at *all* times $S((\forall t)(\exists x)(P_t x))$ and at *some* times $S((\exists x)(\exists t)(P_t x))$, as temporal versions of the principle of plenitude would require.

The principle of sufficient reason itself is characterized by extreme generality. We are tempted to state it in the symbolism just described by saying something like the following "For all entities, x , and for all times, t , if x is P at t , then that (x is P at t) has a sufficient reason for being the case,"

$$(a) (\forall x)(\forall t) [P_t x \supset S(P_t x)]$$

But this is not sufficiently general, as indicated by the fact that Spinoza wants to assert the existence of sufficient reasons for *non-being* as well as being. Thus

$$(b) (\forall x)(\forall t) [-P_t x \supset S(-P_t x)]$$

also captures the idea of the principle, though in a negative form.

But even a conjunction of (a) and (b) would not be sufficiently general. It would have to be generalized to predicates of any degree- n , to be sure, but more important it is not clear that such a conjunc-

tion would cover even so simple a statement as the following, which also falls within the scope of the principle:

$$(c) [(\exists x)(\exists t)P_t x \supset S((\exists x)(\exists t)P_t x)]$$

("If there exists an x which is P at some time t , then that (there exists an x which is P at some time t) has a sufficient reason for being the case"). (c) is not a universally quantified statement like (a) and (b). Whether it can be derived from (a) and (b) is questionable and would depend on what axioms are chosen to define the operator " $S()$ ".⁵

But the most important lesson of (c) for interpreting the principle of sufficient reason is the following: It would appear that the assertibility of *any* formula or sentence, ' A ', in a language in which the principle of sufficient reason were true would imply the assertibility of "that A has a sufficient reason for being the case" in that language. (c) has this form for the closed formula or sentence " $(\exists x)(\exists t)P_t x$." And this suggests in turn that the generality of the principle of sufficient reason is best captured by treating it as an axiom, or thesis, *schema* in the metalanguage of a given language. This is a lead we will follow. A metalinguistic formulation of the principle has the advantage of coming close to Spinoza's own statement of the principle of sufficient reason. For he says that for anything whatever, a sufficient reason *may be assigned* for its existence or non-existence, thus making tacit reference to the ascription or assignment of sufficient reasons by intelligent beings.

One further idea is needed if we want statements of the form of (a) and (b), as well as those of the form of (c), to fall within the scope of the principle. We need the idea of the universal closure " $U(A)$ " of any formula ' A ', defined as follows: If ' A ' contains free individual (including temporal) variables, then ' $U(A)$ ' is the formula ' A ' preceded by universal quantifiers binding all of the free individual variables of ' A '. If ' A ' has no free variables " $U(A)$ " is ' A ' itself.

The principle of sufficient reason is then

$$(P_{SR}) U(A \supset S(A))$$

A principle (applying to any well formed formula, A , in a system characterized as above) which says in effect that anything asserted to be the case (in

⁵ The topic of axioms for " $S()$ " is an interesting one which I do not have the space for here and will consider in another paper. Note, however, that the principle of sufficient reason itself will not count as one of the axioms defining " $S()$ " if it is considered as a synthetic statement (and it is the only non-logical premise needed in the proof).

⁶ Liberties are taken in this and subsequent metalinguistic statements: logical symbols and " $S()$ " are used to stand for themselves. In addition capitals ' A ', ' B ' stand for formulas and later ' T ', ' X ' for temporal and individual variables respectively.

such a system) can be assigned a sufficient reason for being the case.

What is the status of such a principle? Historically, the principle of sufficient reason was not usually regarded as a logical or analytic truth. On some interpretations of the idea of a sufficient reason (for example, the interpretation which makes a sufficient reason a cause) it is at least logically possible that some things do not have sufficient reasons (or causes) for their existence. Those, like Kant, who did regard the principle as necessarily and universally (*a priori*) true in its causal form regarded it as a *synthetic a priori* truth. Note that if plenitude can be derived from (P_{SR}) (as the only non-logical premiss) and cannot be derived without (P_{SR}), as I will claim, and if (P_{SR}) is conceived as a synthetic statement, then the principle of plenitude will also be a synthetic statement, as we would expect.

IV. FROM SUFFICIENT REASON TO PLENITUDE

I want to argue that the principle of plenitude is a special case of (P_{SR}), applying to a subclass of well formed formulae in a system characterized as above, namely, formulae which are negations of well formed formulae having at least one free temporal variable. The simplest example of such formulae, i.e., " $\neg P_t x$," will be used as an example throughout the following proof. In general, the relevant formulae may have any complexity whatever, may involve quantifiers, and may have more than one free temporal variable, but they must be negations and must have at least one free temporal variable. The reasons for these two restrictions will become clear as we proceed. They imply that plenitude will be a principle more limited in scope than sufficient reason.

With this in mind, the derivation of plenitude from (P_{SR}) proceeds as follows: Let " A^* " stand for any formula having at least one free temporal variable, the simplest example being " $P_t x$." We will then be interested in formulae of the kind " $\neg A^*$." The first step of the proof is to show that the principle of sufficient reason, (P_{SR}) $U(A \supset S(A))$, applies to all such formulae " $\neg A^*$." This is simple enough. Since formulae of the kind " A^* " are well formed, formulae of the kind " $\neg A^*$ " will be well formed. Hence, if (P_{SR}) holds for all well formed formulae, then any formulae of the kind

$$(1) \quad \neg A^* \supset S(\neg A^*)$$

will be assertible (for example, " $\neg P_t x \supset S(\neg P_t x)$,"

"If x is not P at time t , then that [x is not P at t] has a sufficient reason for being the case"). (1) is a negative form of the principle of sufficient reason, assigning sufficient reasons for non-existence. Plenitude is related to just such a negative form.

From (1), a number of straightforward logical moves take us to the principle of plenitude. If any formula of the kind (1) may be asserted then by application of the rule of universal generalization, we may assert any formula of the kind

$$(2) \quad (\forall T)[\neg A^* \supset S(\neg A^*)]$$

where " T " stands for a temporal variable free in " A^* " (for example, " $(\forall t)[\neg P_t x \supset S(\neg P_t x)]$," "For all times t , if x is not P at t , then that [x is not P at t] has a sufficient reason for being the case.") Distributing the temporal quantifier, we then have

$$(3) \quad (\forall T)\neg A^* \supset (\forall T)S(\neg A^*)$$

(e.g., " $(\forall t)\neg P_t x \supset (\forall t)S(\neg P_t x)$," "If, for all times t , x is not P at t , then, for all times t , that [x is not P at t] has a sufficient reason for being the case.")

Formulae of the kind (3) apply in just those cases where something is *never* the case (e.g., where x is not P at all times). Such cases have particular relevance to the principle of plenitude. By contraposition, formulae of the kind (3) become formulae of the kind

$$(4) \quad \neg(\forall T)S(\neg A^*) \supset \neg(\forall T)\neg A^*$$

(e.g., " $\neg(\forall t)S(\neg P_t x) \supset \neg(\forall t)\neg P_t x$ ") and by replacement of the expression " $\neg(T)-$ " by " $(\exists T)-$ " (4) becomes

$$(5) \quad \neg(\forall T)S(\neg A^*) \supset (\exists T)A^*$$

(e.g., " $\neg(\forall t)S(\neg P_t x) \supset (\exists t)P_t x$," or "If it is not the case for all times t , that [x is not P at t] has a sufficient reason for being the case, then x is P at some time.")

This reading of the example of (5) suggests that (5) is close to an expression of plenitude. I want to suggest in fact that the principle of plenitude is just the universal closure of (5)

$$(P_{PL}) \quad U[(\neg(\forall T)S(\neg A^*) \supset (\exists T)A^*)]$$

Our example corresponding to this formula would be " $(\forall x)[\neg(\forall t)S(\neg P_t x) \supset (\exists t)P_t x]$ " which says in effect that for all entities, x , if it is not the case that there is a sufficient reason for x 's not being P at all times, then x will be P at some times. Now this, I suggest, is a faithful rendering of the "some times" or "fullness of time" version of plenitude, if we interpret the idea of a genuine possibility in such a way that something is a genuine possibility

if and only if it is *not* the case that there is a sufficient reason for its *non-existence* at *all* times. On this reading, instances of (P_{PL}) say that if something is a genuine possibility, that is, if there is no sufficient reason for its non-existence at all times, then it will exist at some time, and they express the "fullness of time" version of plenitude.

Comments: (i) The proof depends upon interpreting the idea of a *genuine* possibility in terms of the idea of a sufficient reason. There is historical precedent for such an interpretation in Spinoza. He interpreted plenitude in such a way that entities were genuinely possible just in case there was *no* sufficient reason for their *non-existence* (Ethics I, Prop. 11; cf. Lovejoy, 152–153). This is the idea expressed in (P_{PL}) without the references to time. Spinoza related the idea of a sufficient reason in turn to the divine intellect so that there was a sufficient reason for the existence of something if and only if it were conceivable by God. Transferred to plenitude in the manner of the proof, this amounts to interpreting a genuine possibility as a *logical* or *conceptual* possibility and Spinoza did in fact hold the view that all logical or conceptual possibilities must exist or be realized.

Although not explicit in philosophers other than Spinoza, this interpretation of "genuine possibility" in terms of sufficient reason in (P_{PL}) fits a variety of other interpretations of plenitude and provides an interesting perspective on them. On most other interpretations of plenitude, "genuine" possibility is a more restricted notion than mere logical possibility. A more common theological conception than Spinoza's referred sufficient reasons for the existence of contingent entities not only to the divine intellect, but also to the divine will (to the *goodness* of entities or their contribution to the overall perfection of the created universe). The theme, stated for example by Abelard (*Intro. ad Theologiam*, III) "God neither does nor omits to do anything except for some rational and supremely good reason," goes back to Plato's *Timaeus*, was common in medieval thought, and was developed later by Leibniz. In such a tradition not all logical possibilities are "genuinely" possible, but only those which have certain values or contribute to the overall perfection of the created universe. Stated negatively, in the manner of (P_{PL}): if the *non-existence* of something at *all* times would *not* contribute to the overall perfection of the created

universe, then it *will* exist at some time. This is not far removed from Leibniz' restricted version of plenitude which Rescher calls the principle of "existence maximization" in the best of all possible worlds.⁷

Another important restricted account of "genuine possibility" (one we will be exploring in later sections of this paper) results if we interpret the idea of a sufficient reason as the modern idea of a "cause." On this interpretation, a state of affairs has a sufficient reason for being the case if and only if it is caused in the technical sense that there exists a set of antecedent (and/or simultaneous) states of affairs whose joint occurrence is a *sufficient* condition for the occurrence of the state of affairs, *given the laws of nature*. The principle of sufficient reason (P_{SR}) becomes the principle of causality (every event has a cause) on this interpretation and "genuine possibility" in the principle of plenitude (P_{PL}) becomes "physical" or "causal" possibility, a more restricted notion than logical possibility. An occurrence is genuinely or physically possible in this sense if it is not *ruled out* by the laws of nature and antecedent conditions of the universe. Stated more precisely, in the manner of (P_{PL}), a physical possibility is something whose *non-occurrence* is *not* required at *all* times by the laws of nature and antecedent conditions of the universe. Given our present knowledge of the laws of nature, a perpetual motion machine or a particle of ordinary mass traveling faster than the speed of light are not genuinely possible in this sense, though they are logically possible.

The point of introducing these examples is to show that a variety of interpretations of the principles of sufficient reason and plenitude will be compatible with the proof. To each interpretation of the idea of a sufficient reason in (P_{SR}) there will correspond an interpretation of the idea of a genuine possibility in (P_{PL}) and hence a different interpretation of the principle of plenitude. Different interpretations may be subject to criticism and numerous criticisms have been made, for example, against Spinoza's interpretation and the Leibnizian interpretation in terms of the best world. But such criticisms do not affect the relation between the two principles, nor should they be construed as criticisms of the principle of plenitude *per se*, so much as criticisms of what should constitute a sufficient reason for existence.

(ii) The second important implication of the

⁷ Nicholas Rescher, *The Philosophy of Leibniz* (Englewood Cliffs, 1967), pp. 50–51.

derivation concerns the two temporal versions of plenitude discussed in section 2. As I said there, the temporal version of plenitude which can be derived from the principle of sufficient reason is the "some times" or "fullness of time" version of the principle. The role of the temporal quantifiers in the proof shows clearly why the "nature is constantly full" version cannot be derived. It may seem strange that even the "some times" version of plenitude is derivable but we can now see that this is just a particular application of the principle of sufficient reason, an application to cases where there is no reason why something may *never* exist. To describe such cases, the statement of plenitude must be limited to negations of formulae A^* having at least one free temporal variable, which explains the restriction to formulae of this kind.

(iii) The third and most important lesson of the derivation is that plenitude and sufficient reason are not *equivalent* principles; *the derivation cannot be reversed*. Plenitude follows from sufficient reason, but not sufficient reason from plenitude. It would appear that those who intuitively felt that sufficient reason was the more general and fundamental principle were correct. The role of time in the derivation is the critical factor. Step (1) restricts the derivation to formulae with free temporal variables and steps (2) and (3) exploit this restriction. The critical move is from step (2) to (3) which is not reversible. The schema " $((\forall T)[\neg A^* \supset S(\neg A^*)]) \supset ((\forall T)) \neg A^* \supset (\forall T)S(\neg A^*)$ " expresses a logical truth, but its converse does not. When we move from (2) to (3) we move from talk about sufficient reasons for *any* entity existing at *any* time to talk about sufficient reasons for a restricted class of entities which exist at no times or *never* exist.

What is the significance of this? If plenitude follows from sufficient reason, then any world in which the principle of sufficient reason holds is a world in which the principle of plenitude holds also. But if sufficient reason does not follow from plenitude, then there are possible worlds in which plenitude might hold and in which sufficient reason does not hold. This result has interesting consequences for discussions of determinism and chance.

It is generally understood that a world in which the principle of sufficient reason holds on the causal interpretation (i.e., a world in which every

occurrence has a cause) is a deterministic world. It has never been altogether clear whether plenitude also implies a deterministic world. Perhaps the creativity of the divine life is inexhaustible, though its creations are not exactly predictable. On the other hand, if rationalistic philosophers were successful in their attempts to derive the principle of plenitude from sufficient reason that would seem to have settled the matter; plenitude would also imply determinism. Lovejoy seems to have thought that it did and criticizes the principle of plenitude for its deterministic implications in the final pages of his book (331-333). But if the derivation is not reversible, the determinism of the principles of sufficient reason and causality is not *implied* by plenitude. Plenitude will hold in a deterministic world, but does not imply a deterministic world. We should be able to imagine a world in which plenitude holds but sufficient reason and causality do not. I now want to show that there is an example of such a world at hand, abstract, but more than merely imaginary, the world of modern elementary particle physics.

V. PHYSICS AND PLENITUDE

A majority of the "elementary" particles discovered in the last few decades by physicists undergo spontaneous processes of decay, transmuting into particles of different kinds, and researchers have determined which of these decay processes were possible and which were not by focussing on known conservation laws. Thus, for example, at one stage of this research, transmutations which did not violate the known laws of conservation of (mass-) energy, linear and angular momentum, charge and parity, were assumed to be possible and those which violated such laws were assumed to be impossible. If a transformation occurred which violated one of these conservation laws then the universality of the law would be called in question, as in the case of the law of conservation of parity under the impact of the famous experiment of Lee and Yang.⁸

But there is a different kind of result in the history of research into elementary particles which I would like to consider, less startling, but more common than the violation of a known conservation law. The "downfall" of parity was a case of observing events which appeared to be ruled out

⁸ C. N. Yang, *Elementary Particles: A Short History of Some Discoveries in Atomic Physics* (Princeton, 1961). This book, though brief and sketchy, is a useful sourcebook for the historical examples discussed in this section.

by accepted conservation laws. The cases I have in mind are cases in which certain transmutations of elementary particles were *not* observed, even though they were *not* ruled out by known conservation laws.

Consider, for example, the proton. There were various conceivable ways in which it might have decayed into smaller particles without violating the known conservation laws of mass-energy, momentum, charge, etc. (e.g., into a positive electron, a neutrino and antineutrino). But protons do not appear to undergo decay by this or any other process. They are stable particles. The failure to observe spontaneous decay processes involving protons led to the formulation of a new conservation law which prevented such processes from occurring. The law of heavy particle, or baryon, conservation might be stated roughly by saying that in any process involving elementary particles, the number of baryons (a distinct class of particles with half integral spin and comparatively large mass, including protons, neutrons and larger unstable particles) minus the number of antibaryons must remain constant.

Consider the process of reasoning which lies behind the postulation of a conservation law of this kind. Baryon conservation is not an isolated example, for a similar pattern of reasoning lies behind the introduction of other conservation laws in this field, e.g., conservation of leptons or light non-nuclear particles like electrons and muons and some other examples to be cited later. The common pattern seems to be something like the following: If a certain (transmutation) event is not ruled out by known (conservation) laws or antecedent conditions then it should sometimes be observed. Thus, if it is never observed, it is reasonable or plausible to suppose that some unknown conditions or hitherto unknown laws rule it out. Now the first of these two statements is an instance of the principle of plenitude in the form of (P_{PL}) and the second statement is just the contrapositive of the first. Let us restate them more generally in the form of (P_{PL}): If there is *no* sufficient reason for the *non-existence* of a process (spontaneous decay) involving entities of a certain kind (protons) *at all times* (that is to say, on the causal interpretation of sufficient reason, if antecedent condi-

tions and laws of nature do not rule out a spontaneous decay process for protons at all times) then the process will occur at some times, so that, if the process never occurs, conditions and laws must exist which always rule it out.⁹

These statements suggest the *plausibility* of assuming that *some* law or conditions must rule out spontaneous decay processes for protons. Other plausibility considerations and relevant analogies must suggest the further step, that the critical factor is a conservation law of a certain form. Thus the statement of plenitude plays a role analogous to the role played by the principle of causality (i.e., (P_{SR}) on the causal interpretation) in the discovery process. This latter principle suggests the *existence* of explanations of a certain sort, causal (or deductive-nomological explanations), for all occurrences. It does not generate particular explanations, nor the particular forms of explanatory laws, but if accepted as a heuristic principle, it guarantees the existence of explanations of a given kind.

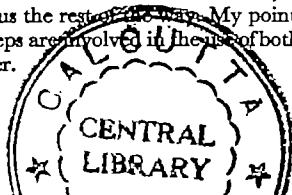
Plenitude functions in a similar way, but with a revealing difference. Plenitude applies only in the *negative* cases where a process of a certain kind is *never* realized and it guarantees the existence of a causal or deductive-nomological explanation only in such cases (i.e., laws which rule out the process at all times), whereas causality is a more general principle guaranteeing a causal or deductive explanation for any occurrence at any time. This is as we should expect if plenitude is a principle more limited in scope than sufficient reason or causality.

The functioning of plenitude here is revealing in another way. The quantum world is not a deterministic world in the sense that the principle of causality (the causal form of (P_{SR})) does not universally hold in it. But if scientists continue to assume that there is a reason or explanation in the form of conservation laws whenever certain processes *never* occur then plenitude continues to be assumed in such a world. We have our example of an indeterministic world in which plenitude continues to hold.

Transmutation processes among elementary particles do not exhaust the possible uses of plenitude in physics. The magnetic monopole provides an example of another sort.¹⁰ Known laws do not

⁹ There is an additional move here which is not brought out, from no transmutation process of this kind has ever been *observed* to no such process has or will *occur*. This is the inductive move (from no observed A 's have been B 's, to no A 's are B 's). After it the principle of plenitude, functioning like the causal principle takes us the rest of the way. My point is that plenitude functions in a way that is analogous to the principle of causality. Inductive steps are involved in the use of both principles.

¹⁰ These examples and the next were suggested to me by Robert Palter.



presently rule out the possibility of such monopoles. It is true that some physicists doubt that magnetic monopoles will be found in nature, while others think they will be found. But most tend to agree to the following: Either magnetic monopoles will be found somewhere in nature or we will gain a deeper understanding of the laws of electrodynamics which will show why such monopoles cannot exist. This disjunctive belief presupposes the principle of plenitude as we have defined it. If laws are not found ruling out magnetic monopoles, then they will exist somewhere, sometime.

One other example comes from a paper by O. Bilaniuk and E. C. George Sudarshan on particles (with imaginary proper mass) whose speed exceeds that of light, the so-called tachyons.¹¹ These authors demonstrate that such superluminal particles are not ruled out by special relativity as once thought and could be emitted by some as yet undetermined process. They then add: "If their [i.e. tachyons] existence would not lead to any contradictions [with laws of physics] one should be looking for them. There is an unwritten precept in modern physics, often facetiously referred to as Gell-Mann's totalitarian principle which states that in physics 'anything which is not prohibited is compulsory.' Guided by this sort of argument we have made a number of remarkable discoveries from neutrinos to radio galaxies." (p. 44)

We should no longer wonder about the source of this "unwritten precept" of modern physics which these authors call Gell-Mann's totalitarian principle. It is a very old principle called plenitude and it is logically related to a more familiar principle called causality.

VI. PRINCIPLES OF EXPLANATION

What reason do we have for continuing to hold onto the principle of plenitude in the quantum world where the causal principle no longer holds? This is an interesting question related to some important problems of scientific methodology. While I cannot supply a completely satisfactory answer, the following comments may throw some light on the general functioning of heuristic principles in explanation. I want to suggest that something like the following heuristic rule ought to, and does, function in scientific contexts regarding principles of explanation like causality and plenitude: "Assume the strongest principle of explanation

in one's search for explanatory hypotheses which is not falsified by the empirical evidence." We have noted that causality is a stronger principle of explanation than plenitude, calling for explanation of a wider class of phenomena. In fact, since deductive explanations can be assumed to be stronger than statistical ones, causality can be assumed to be the strongest principle of explanation in scientific contexts. In other words, other things being equal, the principle of causality should be assumed first and until such time as evidence fails to support it.

Admittedly there is a problem about empirical falsification for principles like causality and plenitude. But if modern physics provides any clues, there are ways in which principles of explanation, like the principle of causality, can be *indirectly* falsified. The principle may be incompatible with the best available theory (in this case, quantum theory), where the notion of the "best available theory" is defined by the usual criteria, as the theory accounting for the widest range of phenomena with the greatest accuracy, and the simplest theory doing so (which is not to say that there are no methodological problems about how these criteria are to be defined and applied).

But the same evidence which supports a theory that is incompatible with the causal principle need not also rule out the weaker principle of plenitude. Quantum theory rules out the principle that a causal or deductive nomological explanation must be available for every particular occurrence, but does not rule out the principle that a causal or deductive nomological explanation must be available whenever processes of certain kinds fail to occur *at all times*. What I am arguing is that, in such a case, we do not need a separate justification of plenitude. Our continuing to hold plenitude in face of the denial of causality follows the rule: "If a principle of explanation is (indirectly) falsified by the empirical evidence, fall back on those weaker principles of explanation not yet falsified by the same evidence."

In other words, we should be asking a different sort of question than the one originally asked, not the question "Why continue to hold the principle of plenitude if the principle of sufficient reason or causality is false?", but the (Popperian) question "Given that sufficient reason or causality is assumed to be false, what *further* reasons would lead us to regard plenitude as also being falsified?"

¹¹ O-M. Bilaniuk and E. C. George Sudarshan, "Particles Beyond the Light Barrier," *Physics Today*, vol. 22 (1969), pp. 43-52.

What kind of a physical theory would provide indirect evidence for the falsity of plenitude, if it were accepted as the best available theory? It would be a theory which implied that some kinds of events which never occurred, and never would occur, could nevertheless not be ruled out in principle by any set of laws of nature *cum* antecedent conditions. A strange theory indeed! But before we say there could be no such theory, we should reflect on the fact that even the best physicists of the 19th century could not have imagined what a theory would look like which would rule out *in*

principle the universal application of the causal principle. And this implication of quantum physics was difficult to believe for some of the greatest minds, like Einstein's, even after the theory had been formulated. If anyone ever comes up with a theory incompatible with plenitude, philosophers as well as scientists are likely to learn a good deal from it, just as they have learned from quantum theory not to repeat Kant's mistake and assume that the principle of causality is a necessary condition for experiential knowledge of natural phenomena.¹²

The University of Texas at Austin

Received January 22, 1975

¹² I would like to thank the following persons who either discussed issues of this paper with me or read an earlier draft or both: Gerd Buchdahl, Brian Cooney, Charles Hartshorne, Bernard Katz, Norman Martin and Robert Palter.

IV. THE EXISTENCE OF MENTAL OBJECTS

FRANK JACKSON

I

THE mental objects I will be concerned with in this paper are such things as pains, itches and throbs—that is, the bodily sensations; and such things as after-images and mirages—that is, the visual hallucinations. There is a very widespread view that, while there may be things like the *having* of bodily sensations and the *experiencing* of after-images, there are, strictly speaking, no such things as bodily sensations and after-images. What exists includes the experiencing of pains and after-images, but not the pains and after-images themselves.

I will argue that there are considerable problems facing this denial of mental objects, and to this extent support that there are mental objects and the associated act-object account of having sensations and after-images. (The view that sensations and images exist, is just that. It is not that they exist *independently* of sentient creatures. It is an open question whether everything (that exists) exists independently, one to be settled by looking at cases—including ours of sensations and images.)

II

We talk as if there were mental objects: "*There is a pain in my foot,*" "*This after-image is brighter than that one,*" and so on.

This settles nothing as it stands. We once talked as if there were demons, and we now often talk about the average family or the next waltz; yet there are no demons, average families or waltzes. These three examples illustrate three ways to show that there are no *As*. We were wrong about demons, because certain statements we took to be true turned out false (epilepsy is *not* caused by demons). There is no average family because statements that appear to be about this family can be given a reductive style of analysis in terms of the many non-average families that there are. The case with waltzes is different. No doubt, "The waltz is about to start" can be given a reductive analysis in which "waltz" does not appear. But it is sufficient to observe that "The waltz is about to start" can be construed as "People are about to start waltzing." We don't

need a full-scale analysis of waltzing, we only need to know enough about the meaning of "waltz" to know that statements putatively about waltzes can be, and are best, re-cast as about people waltzing.

Likewise, there are three ways we might seek to show that there are no mental objects: by showing that all statements of the form "*S has a pain (itch, after-image, etc.)*" are false; by producing a reductive analysis of such statements, for example, of a behaviourist or topic-neutral kind, which eliminates the relevant psychological terms; or, finally, by offering a partial analysis (a recasting which better displays logical form) of these psychological statements, and which, while not eliminating all mentalist vocabulary, shows that these statements are not really about mental objects.

Of the first strategy, I will just say that I am sure it is mistaken, but do not know how to prove that it is. For I do not know of any premisses which are more obvious than that it is sometimes true that we are in pain, having a red after-image, and so on, from which a proof might be constructed. The second strategy has been much discussed in connection with the translation versions of materialism advanced by J. J. C. Smart and D. M. Armstrong. I don't find the analyses they give of psychological statements plausible for reasons that are familiar and which will be taken as read.

I will concentrate on the third strategy in this paper. I am sure the popularity of the denial of mental objects is due to the belief that it can be sustained by a relatively simple recasting of visual image and sensation statements without recourse to either a wholesale rejection of their truth or a full-scale, reductive analysis.

III

With one exception, our discussion of the third strategy applies equally to bodily sensations and visual images. The exception arises from Bruce Aune's suggestion that sensations can be regarded as relations between persons and their bodies. His ground for this is that "'I have a pain in my arm' . . . may be rephrased as, 'My arm pains me' or 'My

arm hurts'.¹ But it is impossible for a relation to hold in the absence of its relata. Hence, if sensation statements essentially related persons to parts of their body, they could not be true in the absence of appropriate parts of the body. But the phantom limb phenomenon shows that they can be; and a statement that I have a pain "in a phantom limb" cannot be regarded by Aune as relating me to my phantom limb, because either phantom limbs don't exist or else they are a species of mental object (part of the "body-image"). Moreover, even if—as has sometimes been maintained—a pain in my phantom limb is really in my stump, the statement cannot be rephrased as "My stump pains me," because this is the translation of a quite different statement according to the view in question, namely "I have a pain in my stump."

IV

I now turn to examples to strategy three which are equally applicable to sensations and images. I will start by talking in terms of the sensation case and switch later, for variety, to the image case.

The following passage from Thomas Nagel's paper, "Physicalism," makes a convenient starting point.

... we may regard the ascription of properties to a sensation simply as part of the specification of a psychological state's being ascribed to the person. When we assert that a person has a sensation of a certain description *B*, this is not to be taken as asserting that there exist an *x* and a *y* such that *x* is a person and *y* is a sensation and *B*(*y*) and *x* has *y*. Rather we are to take it as asserting the existence of only one thing, *x*, such that *x* is a person, and moreover *C*(*x*), where *C* is the attribute "has a sensation of description *B*."²

Nagel's general idea is to switch from predicates on or descriptions of sensations, to predicates on or descriptions of persons: strictly, nothing is painful, but many things are persons with painful sensations. But how does the switch he recommends dispose of sensations?

Every description of my brother can be transposed to a description of me without meaning loss: "My brother is tall," for instance, goes to "I have a tall brother." But the possibility of switching from "is tall" as a predicate on my brother to "has a tall brother" as a predicate on me is irrelevant to the

question of my brother's existence. What matters is how we ought to understand the predicate "has a tall brother"; the answer in this case being that it is to be understood as formed from a relation by filling an argument place with a singular term. Likewise, what is crucial for whether sensations exist, is not just that a statement like "My pain is severe" can be rendered as "I have a severe pain"; but whether or not it can be so rendered with the predicate "has a severe pain" understood other than as containing "a pain" functioning as a singular term filling an argument place in the relation "*x* has *y*."

V

Can we, then, view the semantic structure of "has a painful sensation" and the like so as not to commit ourselves to there being painful sensations?

The simplest such view would be the view which sees nothing; the view, that is, that sensation predicates on persons have no semantic structure at all; they have meaning only as wholes; a view we might express by writing "has-a-painful-sensation" or "hasapainfulsensation."

This is holism gone mad. It is perfectly obvious, for example, that the meanings of "has a burning pain" and "has a burning itch" are related, and that this relation is a function of the common term "burning": if they did not have this term in common, the relation would not be the same. But to concede that the appearance of "burning" in *both* is semantically significant is to concede that its appearance in *either* is, which is to concede that the predicates are semantically structured.

Given the unacceptability of this no-structure theory, what theories are open to one who denies that there are mental objects? Just two seem to have either currency or plausibility: the first I will call the state theory, the second the adverbial theory.

VI

Those who deny that there are mental objects commonly make considerable use of terms like "state" and "condition." For instance, Nagel urges that to say that a person has a pain "describes a *condition* of one entity, the person," and again "we may regard the ascription of properties to a

¹ *Knowledge, Mind, and Nature* (New York, 1967), p. 130.

² *The Philosophical Review*, vol. 74 (1965), p. 342.

sensation simply as part of the specification of a psychological *state's* being ascribed to the person."³

This suggests a state theory of the following kind. "I have a throbbing, painful sensation in my knee" does not relate me to my sensation, saying of it that it is throbbing, painful and in my knee; rather it is about me and a state (or condition) of mine: it says that I am in a throbbing, painful, in my knee sensation-state. The sensing, not the sensation has the properties.

Does this really achieve anything? On the act-object theory, "I have a painful sensation" is explicated as " $(\exists x)[I \text{ have } x \ \& \ x \text{ is painful} \ \& \ x \text{ is a sensation}]$," while on the state theory we get " $(\exists x)[I \text{ am in } x \ \& \ x \text{ is a sensation state} \ \& \ x \text{ is painful}]$." The gain appears to be verbal rather than ontological.

The state theorist has some sort of reply if he distinguishes *unitary* states from *relational* states. A unitary state of a person is a state of that person not essentially involving anything over and above that person. My being happy, to take a psychological example, and my being warm, to take a physical example, are both unitary states of mine. There is a natural, if philosophically difficult, sense in which my being happy or being warm at some time are not things over and above and distinct from me at that time. This is essentially linked to the connection between the counting principles for persons and their unitary states: for a given person at a time, there cannot be more than one unitary state of a given kind. For instance, there cannot be more cases of persons being warm (happy) in a room than there are persons. If this were not so, if there might be two unitary states of a given kind for one person at one time, there would be no sense to the claim that such states were, in some substantial sense, nothing over and above the one subject in these states. Therefore, if the state theorist insists that the sensation states—the sensings—are unitary, rather than relational states like being happy *at* or being warmed *by*, it appears he has a theory distinct from the act-object theory.

He also has a theory exposed to two serious objections. I will call the first, the many property objection, and the second, the complement objection.

VII

The many property objection arises from the fact that we ascribe many things to our sensations: a

sensation may be painful *and* burning *and* in the foot. How can a state theorist handle this?

The state theorist transcribes "I have an *F* sensation" as "I am in an *F* sensation-state." Hence, the obvious account for him to give of "I have a sensation which is *F* and *G*" is "I am in a sensation-state which is *F* and *G*."

This account faces a decisive difficulty. Suppose I have a sensation which is *F* and a sensation which is *G*, then, on the state theory, I am in a sensation-state which is *F* and in one which is *G*. But there may at a given time be only one such unitary state for a given person; therefore, I am in a sensation-state which is *F* and *G*. But the latter is the preferred account of "I have a sensation which is *F* and *G*." Hence, we have "I have a sensation which is *F* and sensation which is *G*" entailing "I have a sensation which is *F* and *G*," which is quite wrong. For I may have one sensation which is *F* and, at the same time, another which is *G*. That is, having, for example, a burning, painful sensation is being conflated with having a burning sensation and a painful sensation.

The state theorist must, therefore, give a different account of statements that a particular sensation is *F* and *G* and. . . . It turns out that the various possibilities are essentially the same as those that arise in the discussion of the corresponding objection to the adverbial theory, so I will postpone the matter until then.

VIII

The complement objection to the state theory is a special case of a general way of showing that some term does not qualify a given thing, the way which proceeds by showing that if the term did, so might its complement.

The view that truth is a property of sentence *types* may be refuted by noting that "is true" and its complement "is not true" may apply to one and the same sentence type depending on the meanings given to the constituent terms of that type. A second example is the view that "school-age" in "I have a school-age child" qualifies having a child rather than the child; that is, that being of school-age is, strictly, a property of having children rather than of children. But "I have a school-age child" and "I have a non-school-age child" are both true. And nothing, including the having of children, can be both *F* and non-*F*, hence it is the children which are or are not of school-age.

³ *Ibid.*, p. 342. My emphasis.

Likewise, it may be the case that "I have a painful sensation" and "I have a non-painful sensation" are both true at the one time, hence we cannot construe being painful or not as a characteristic of the having of the pain rather than the pain. For if we did, we would have a state being both *F* and non-*F*. As with the first objection, the possible replies are essentially the same as those to the corresponding objection to the adverbial theory, and will thus be discussed then.

(A digression: the possibility of having different kinds of sensations at the one time also seems to me to undermine Descartes' suggestion that we view sensations on the model of impressions in wax: "I allow only so much difference between the soul and its ideas as there is between a piece of wax and the various shapes it can assume."⁴ But it is impossible for one thing to be two different shapes at the same time; hence the analogy fails at a crucial point.)

IX

The most widely canvassed theory which denies that there are mental objects is the so-called adverbial theory.⁵ Our discussion will be couched principally in terms of visual images (after-images, for example) rather than bodily sensations.

This theory utilises the fact that after-images, sensations, and the like, cannot exist when not sensed by some person (sentient creature), in order to reconstrue statements which purport to be about sensations, after-images and so on, as being about the way or mode in which some person is sensing. Thus, "I have a red after-image" becomes "I sense red-ly," and "I have a pain" becomes "I sense painfully." A parallel is often drawn with the elimination of talk about smiles in favour of talk about the manner of smiling; as in the recasting of "Mary wore a seductive smile" as "Mary smiled seductively."

My two objections parallel the two brought against the state theory.

X

Our statements about visual images are not just to the effect that an image is red, or square, or whatever; they are also to the effect that an image

is red *and* square *and*. . . . As with the state theory, I will refer to this as the many property problem.

Adverbial theorists have been rather reticent about how they handle this problem. It is clear that their view is that to have an image which is *F*, is to sense *F*-ly—the attribute, *F*, goes to the mode or manner, *F*-ly. But it is not clear just what account would be offered of having an image which is *F* and *G*. Do both of the (in their view, apparent) attributes go to separate modes, so that to have an image which is *F* and *G* is analysed as sensing *F*-ly and *G*-ly; or do we have a new, compound mode, *F*-*G*-ly? It seems to me that both of these answers, and the variants on them, face substantial difficulties.

Suppose having an *F*, *G* image is analysed as sensing *F*-ly and *G*-ly.⁶ This conjunctive style of answer has the advantage of explaining the entailment from "I have a red, square image" to "I have a red image"; for it will correspond to the entailment from "I sense red-ly and square-ly" to "I sense red-ly."

But if it is adopted, it will be impossible for the adverbial theorist to distinguish the two very different states of affairs of having a red, square after-image at the same time as having a green, round one, from that of having a green, square after-image at the same time as having a red, round one; because both will have to be accounted the same, namely, as sensing red-ly and round-ly and square-ly and green-ly. In essence, the point is that we must be able to distinguish the statements: "I have an *F* and a *G* image," and "I have an *F*, *G* image."

In discussion it has been suggested to me that the adverbial theorist might have recourse to the point that when I have a red, square after-image at the same time as a green, round one, they must (as we say) be in different places in my visual field: the red one will be, for instance, to the left of the green one. But how can this help the adverbial theorist? For "I have a red after-image to the left of a green one" raises the same problem; namely, that it cannot be analysed conjunctively as "I sense red-ly and to-the-left-ly and green-ly"; for that is equivalent to "I sense green-ly and to-the-left-ly and red-ly" which would be the analysis of "I have a green after-image to the left of a red one."

⁴ *Descartes Philosophical Writings*, tr. by Elizabeth Anscombe and P. T. Geach (Nelson, 1954), p. 288.

⁵ Advanced in, e.g., C. J. Ducasse, *Nature, Mind, and Death* (Illinois, 1951); R. M. Chisholm, *Perceiving* (Ithaca, 1957); Aune, *op. cit.*

⁶ This is the obvious reading of Ducasse, *op. cit.*, ch. 13, §22.

Perhaps the thought is that "red" and "square" when applied to images are *incomplete*, they demand suppletion with a term indicating location in a visual field. But this can't be right. I can know perfectly well what saying someone has a red or a square after-image means without having any idea at all of its location. Moreover, it is evidently not possible to give an exhaustive list, $p_1, \dots, p_n$ of all the parts of a person's visual field that might be occupied by one of his after-images. So that "I have a red after-image" cannot be analysed as "I sense red- p_1 -ly or $\dots$ or red- p_n -ly." The best that can be done is " $(\exists x)[x$ is a part of my visual field and I sense red- x -ly]," which—leaving aside the question of interpreting quantification into adverbial modification—commits the adverbial theorist to the existence of a species of mental object, namely, parts of visual fields, and so undermines the whole rationale behind his theory.

XI

An alternative approach would be to translate "I have an F, G image" as "I sense F -ly G -ly," and model the latter on "He wrote astonishingly slowly": " F -ly" being taken to modify " G -ly" rather than "sense," just as "astonishingly" modifies "slowly" rather than "spoke" in the model.

But the two essential features of "He wrote astonishingly slowly" are, first, that it is evidently not equivalent to "He wrote slowly astonishingly," and, second, that it does not entail "He wrote astonishingly": it is these two features which lead us to distinguish it sharply from "He wrote astonishingly and slowly."

The adverbial theorist, however, cannot attribute either feature to his translations. He must allow that "I sense red-ly square-ly" is equivalent to "I sense square-ly red-ly," for they are, respectively, his translations of the evidently equivalent "I have a red, square image" and "I have a square, red image." He must also allow that "I sense red-ly square-ly" entails "I sense red-ly," for "I have a red, square image" entails "I have a red image." Having a red, square image is a special case of having a red image, while writing astonishingly slowly is not a special case of writing astonishingly (just as writing very quickly is not a special case of writing very).

XII

When talking of having a red, triangular sense-impression, Wilfrid Sellars talks of sensing red-triangular-ly.⁷ The hyphenation suggests an interpretation according to which red-triangular-ly is not a mode of sensing having red-ly as a component; it is, rather, a quite new mode of sensing; and so, the meaning of "red-triangular-ly" is not to be viewed as being built out of independently semantically significant components like "red" and "triangular"; and likewise for "green-square-ly," etc.⁸

Put thus, this view obviously faces the difficulty just considered. Having a red, triangular after-image is a special case of having a red after-image, hence any adverbial theorist must treat sensing red-triangular-ly as a special case of sensing red-ly. But sensing red-triangular-ly fails to have sensing red-ly as even a component on this view.

It might, however, be refined. In discussions of the step from "This is a horse's head" to "This is a head," it is sometimes suggested that the latter should be read as "This is a head of something." In similar vein, it might be suggested that "I have a red image" should be expanded to "I have a red image of some shape," and consequently its adverbial translation should be expanded to "I sense red-some-shape-ly." On this view, red-ly isn't a mode of sensing at all. The modes of sensing are red-triangular-ly, green-round-ly, and so on; and sensing red-ly is to be understood as sensing red-squarely or red-round-ly or red- $\dots$ -ly.

There are two serious difficulties facing this suggestion (apart from that of giving a precise construal of the dots). The first is that the modification appears to undermine the adverbial theorist's claim to be offering a philosophically perspicuous account of after-images. When I have a red, square after-image, the redness and the squareness appear as discriminable elements in my experience; and hence elements that it is desirable to have reflected in distinct elements of any offered analysis. But, on the modification in question, having a red, square after-image is accounted as sensing red-square-ly, where the hyphenation indicates that this mode of sensing is not to be further broken up into distinct elements. Indeed, on this view, someone who remarks on the common feature in having red, square

⁷ In, e.g., *Science and Metaphysics* (London, 1968). Sellars ties his view to a topic-neutral analysis of the kind I have said (without arguing) is implausible.

⁸ Such a view is explicitly advanced by George Pitcher, "Minds and Ideas in Berkeley," *American Philosophical Quarterly*, vol. 6 (1969), pp. 198-207.

after-image and having a red, round after-image is making a plain mistake. But, far from being a plain mistake, the remark looks like an evident truth.

The second objection derives from the point that there are indefinitely many things that may be said about one's images. An after-image may be red, red and square, red and square and fuzzy at the edges, red and square and fuzzy at the edges and to the left of a blue after-image, and so on. Now consider how the adverbial theory should handle "I have a red, square, fuzzy after-image." It cannot analyse this as "I sense red-square-ly and fuzz-ly," for the same reason the conjunctive account had to be rejected. In brief, such a treatment would conflate "I have a red, square, fuzzy after-image" with "I have a red, square after-image and a fuzzy after-image." Should the theory then abandon the view that red-square-ly is a fundamental mode of sensing, and adopt the view that red-square-fuzzy-ly is a fundamental mode of sensing? On this further modification, "I have a red, square fuzzy after-image" would go to "I sense red-square-fuzzy-ly"; and "I have a red, square after-image" would be analysed as, roughly, "I sense red-square-fuzzy-ly or I sense red-square-sharp-ly." Thus, on this further modification, red-square-ly, green-round-ly, and so on, are no longer modes of sensing; rather red-square-fuzzy-ly, green-round-sharp-ly, and so on, will be the various ways of sensing.

However, in view of the point this objection started with, this process of modification will continue without end. For any n that the adverbial theorist offers an analysis of "I have an $F_1, \dots, F_n$ after-image" as "I sense $F_1\text{-} \dots \text{-} F_n\text{-ly}$," he can be challenged for his analysis of "I have an $F_1, \dots, F_{n+1}$ after-image"; and so, for the reasons above, be forced to abandon $F_1\text{-} \dots \text{-} F_n\text{-ly}$ in favour of $F_1\text{-} \dots \text{-} F_n\text{-}F_{n+1}\text{-ly}$ as a basic mode of sensing. This means that the adverbial theorist cannot ever give even a single example of a basic mode of sensing, and thus cannot ever complete even one of his adverbial analyses; and even if he could, would, moreover, end up with a theory no better than the no-structure one rejected earlier.

I suspect that some adverbial theorists who have written down expressions like "red-square-ly," have meant by the hyphenation no more than that mode of sensing associated with what we normally, and in their view misleadingly, call having a red, square after-image. But this is not to give us a theory we can oppose to the act-object theory, it is merely to express the hope that such a theory may be

forth-coming. It is not to argue or show that we can do without mental objects, it is just to say that we can; for the central question of how to interpret the hyphenation is left answered except for a reference to the very theory being denied.

XIII

My second objection to the adverbial theory is the complement objection transferred from the state theory to the adverbial. Just as it is not possible for something to be F and non- F at the same time, it is not possible for a person at a given time to V both $F\text{-ly}$ and non- $F\text{-ly}$. I can sing badly easily enough, but I cannot sing both well and badly at the same time; I can run quickly, but not both quickly and slowly; and I can inspect carefully, but not both carefully and carelessly; and so on and so forth.

Therefore, to have an image which is F cannot be to sense $F\text{-ly}$; for it is manifestly possible to have an image which is F at the same time as one which is non- F : I may have a red and a green after-image at the same time, or a square and a round one at the same time; while it is not possible to sense $F\text{-ly}$ and non- $F\text{-ly}$ at the same time.

The only reply which appears to have any real plausibility here is to urge that, though one cannot V both $F\text{-ly}$ and non- $F\text{-ly}$ at a given time, one can $V F\text{-ly}$ with respect to A and V non- $F\text{-ly}$ with respect to B : I can, during a concerto, listen happily to the strings and unhappily to the piano. And that when I have a red and a non-red after-image together, I am sensing red-ly with respect to one thing and non-red-ly with respect to another. But what are these things with respect to which I am sensing, for there need, of course, be no appropriate physical thing in the offing? It is hard to see what they could be other than the mental objects of the act-object theory.

XIV

Two matters in conclusion.

(i) Terence Parsons claims that

(1) John wrote painstakingly and illegibly.

and

(2) John wrote painstakingly and John wrote illegibly.

are not equivalent because—though (1) entails (2)—(2) does not entail (1). He gives two cases

which he claims show (2) may be true when (1) is false:

... if there were two separate past occasions on which John wrote, on one of which he wrote painstakingly, and on the other of which he wrote illegibly, but no past occasion on which he did both at once ... Also if on one and the same occasion he wrote painstakingly with one hand and illegibly with the other.⁹

This might appear to threaten my arguments in two ways. First, in my discussion of the conjunctive reply in §10 I was clearly working under the general assumption that there is no significant distinction between a statement like (1) and the corresponding statement like (2). Second, Parson's second case where (2) may be true while (1) is false, could easily be modified to threaten the principle that one cannot *V*-ly and non-*V*-ly at the same time, and so my discussion of complementation *vis a vis* the adverbial theory. The modification would be to consider a case where John wrote illegibly with his left hand while writing legibly with his right; would he then be writing legibly and illegibly at the same time?

The threat, however, is more apparent than real. This is obvious in the first case Parsons gives, because it involves considering *different* times of writing; and our discussion of the conjunctive reply involved just *one* time—we noted the possibility of having different visual images at the *same* time. In short, it is sufficient for us if "John is writing painstakingly and illegibly" is equivalent to "John is writing painstakingly and John is writing illegibly," and the first case does not threaten this equivalence.

With the second case we must remember that we are dealing with something that can be judged both overall and in a particular aspect. Normally, when we say that Jones wrote illegibly, we mean that overall the writing was illegible, not that every word was illegible (likewise, a speech may be impressive without every part of it being impressive). In this sense, "Jones wrote illegibly with one hand (his left, say)," does not entail that Jones wrote illegibly, for most of the writing may have been with his right hand in elegant copperplate; and similarly for "painstakingly." And in this sense Parsons will be right that it is possible that Jones wrote painstakingly with one hand and

illegibly with the other without (1) being true, but equally this is possible without (2) being true, so the case fails to establish that (2) may be true without (1) being true.

On the other hand, if we take "Jones wrote illegibly" to count as true if any part or aspect of Jones' writing was illegible, and likewise for "painstakingly"; then if Jones wrote painstakingly with one hand and illegibly with the other, (2) must be true, but so will (1); and so there is still no case for denying that (2) entails (1).

Parallel remarks apply to the possibility of writing illegibly and legibly. It is not, in the overall sense, possible to write, on a given occasion, both legibly and illegibly (though it is possible to write in manner which deserves neither epithet). It is possible that one aspect of one's writing be legible and *another* be illegible. But we noted in §13 that the possibility of *V*-ing *F*-ly with respect to *A* while *V*-ing non-*F*-ly with respect to *B* is of no use to the adverbial theory; for the only plausible candidates for *A* and *B* in the sensing case are mental objects.

In general, whether or not one agrees with my discussion of (1) and (2), there is little comfort for the adverbial theorist in Parsons' remarks. The case for distinguishing (1) and (2) rests heavily on there being something *more* involved than just the person (John's hand as well as John); and the adverbial theorist's aim is to effect an ontic reduction to the person alone in his account of sensing.

(ii) The second matter concerns the "predicate modifier" formal semantics for adverbs given by Parsons. By contrast with Davidson's event-predicate treatment,¹⁰ these semantics view adverbs as functions on predicates; and it might be thought that they could be appealed to by the adverbial theorist to elucidate "green-triangularly" and so on in a way which acknowledged structure without facing the problems of the conjunctive treatment.

The "predicate-modifier" theorist must, however, see a certain *intensionality* in *all* adverbs. "*X* senses" and "*x* breathes" are (we may suppose) co-extensional. But John doesn't breathe slowly if and only if he senses slowly, and the adverbial theorist won't allow that he breathes red-ly when he senses red-ly. Without going into the details,¹¹ this means that possible worlds (and beings) other

⁹ "Some Problems Concerning the Logic of Grammatical Modifiers," *Synthese*, vol. 21 (1970), p. 131.

¹⁰ In "The Logical Form of Action Sentences," *Logic of Action and Preference*, ed. by Nicholas Rescher (Pittsburgh, 1967). Adoption of this treatment would make the adverbial and state theories virtually indistinguishable.

¹¹ But see David Lewis, "General Semantics," *Synthese*, vol. 22 (1970), esp. p. 28.

than the actual must be invoked in predicate-modifier semantics. Hence, they achieve nothing for the adverbial theory. Perhaps (perhaps) we need possibilities for the elucidation of modal statements, but "I have a pain" and "I have a red

image" are statements about the actual world, if any are. Moreover, appeal to possibilities would make a mockery of any claim of the adverbial theory to greater ontological economy than the act-object theory.¹²

La Trobe University

Received October 25, 1974

¹² I am much indebted to Tim Oakley, Tom Richards and Barry Taylor for convincing me of the inadequacies of an earlier treatment of §10, 11, 12 titled "On The Adverbial Analysis of Visual Experience," *Metaphilosophy*, vol. 6 (1975), and to Keith Campbell and Chris Mortensen for comments on earlier drafts of this paper.

V. THE STATUS OF RELIGIOUS BELIEFS

FRANK B. DILLEY

IN recent decades there has been much discussion as to whether it is possible for religious belief-systems to live in peaceful coexistence with naturalistic or empiricist doctrines of factual truth. This effort at rapprochement has taken two forms, the reduction of religious reference to the realm of practical commitments and attitudes, and the attempt to leave the realm of factual truth intact for the empiricist while staking out a realm of religious truth which is gained by methods other than scientific. While these approaches are radically different in some respects, they are united by the abandonment of any claim to challenge naturalism on the fields of reasoned metaphysical inquiry. Both agree that the path of reasoned inquiry leads to naturalism. One response is to accommodate religious meaning to a naturalistic framework while the other is to overthrow that framework on the basis of truths of revelation which must be accepted on authority. The attempt to develop a comprehensive metaphysical outlook which fits rational and scientific knowledge into a religious framework has been abandoned by both.

It should be quite evident that religious belief-systems cannot be grounded by scientific tests for factuality as interpreted by empiricists. The options, given empiricist readings of scientific method, are to justify an hypothesis in terms of direct appeal to sense experience or to justify an hypothesis about unobservables in terms of a prediction of observable events which cannot be predicted on any alternative hypothesis.

Plainly beliefs in supernatural beings cannot be justified in either of these ways, given the kind of world that we inhabit and the kinds of things that do occur in that world. It is possible to imagine alternative worlds, or sets of events which might (but do not) occur in this world, which would constitute validation of religious belief-claims. However in this world the gods do not appear in a publicly verifiable form high in the sky, giving loud predictions and performing miracles in accordance with those predictions, hence religious

belief-claims cannot be established by scientific criteria. Some would say that this is because there are not supernatural beings, but others would say that this is because the tests proposed by science are such that our kinds of supernatural beings are excluded methodologically.

The direct test, observation, rules out validation of God-claims because God, by hypothesis, is not observable by ordinary sense experience. It is sometimes claimed that religious experiences give people access to the Divine, but given the rules of scientific method the claim that the objects of that experience have objective existence cannot be justified. Suppose that it could be shown that under conditions *c* every observer has an experience of the divine (being, consciousness, bliss). Since the object of that experience cannot be located spatio-temporally, the lawful occurrence of religious experience would no more be acceptable as a demonstration of the existence of God than would the fact that everyone might experience pink elephants under conditions *d*. Presumably hallucinations are lawful and will occur uniformly when the proper causal conditions occur.

Religious experiences will be given a subjective interpretation in the absence of any locatable spatio-temporal object of that experience. There are after all, experiences which are demonstrably hallucinatory (optical illusions, pink elephants) and it is a rule of scientific inquiry that a solution which is more complicated should be rejected if an adequate simpler one is available. Cornford and Lehrer say:

Given all this evidence, it is reasonable to conclude that many mystical experiences have natural causes. We should be careful, however, about inferring from this that all mystical experiences can be given naturalistic, scientific explanations. Nevertheless, in the face of such evidence, there is little justification for holding that some mystical experiences are not scientifically explainable. We can conclude, therefore, that there is no reason to postulate that some supernatural force is the cause of religious mystical experiences.¹

¹James W. Cornford and Keith Lehrer, *Philosophical Problems and Arguments: An Introduction* (Second Edition: New York, 1974), p. 344.

Although in the preceding passage Cornford and Lehrer allow for the possibility that there might actually be supernatural causes, they conclude that it would be unjustifiable, methodologically, to postulate supernatural causes. Since there are some cases where the experiences can be explained by naturalistic causes, parsimony requires that naturalistic causes be provided for all the rest.

Religious believers claim that there are supernatural causes of all events, and in addition many religious believers hold that there are events which are caused by God in a special miraculous sense. Many religious believers think that God has sent special messengers or takes human form himself, or has caused miraculous healings to occur as clear evidence of his existence for those who will recognize the miracles. However most of these events are of the sort to which naturalistic causes could be attributed. Others can be disregarded entirely as illusory because of lack of substantiating corroboration. Those corroborated by large groups of people have to be weighed against the evidence of all our other non-miraculous experiences (the Humean rule). By this test, given the kinds of events that occur, no postulation of supernatural causes is ever warranted. These rules are plainly biasing in that supernatural causes, if they did occur, would be denied because of the method. To adopt the methodology is to prevent our ever knowing about or recognizing the existence of whatever supernatural causes there might be.

The prediction option cannot be exercised because any event of the type that religious believers normally expect to occur can be given naturalistic explanation, as is proved by the fact that naturalists have given such explanations to whatever divinely caused events there may have been in the past. It is not as though we were faced with two theories of light, able to devise experiments which would fit one theory but not the other, because whatever occurs will be made to fit each of the theories. The one exception, predictions about rebirth into an order of existence in direct communion with God, is of no use now in this world since you must leave this world to have the experience.

One of the ways in which critics attempt to discredit theism is to pose the question, does theism say anything about the world other than that which non-theisms say? If so, supposedly, there are grounds for taking theism seriously, if not, then it

can be dismissed as a live option for rational people. Renford Bambrough asks the question: "Is it conceivable that God should exist and yet that everything else should remain exactly the same as if he did not exist? Is it conceivable that God should not exist, and yet that everything else should remain the same as if he did exist?"² He goes on to claim that transcendentalist theology has given no adequate answer to this challenge.

If one takes the word "adequate" at its face value, one can only be baffled by the question, but once one comes to understand that the word is a portmanteau for "adequate to meet Bambrough's criteria" then the question makes sense, and his claim that transcendentalist theology gives no adequate answer is correct. It is correct because it is impossible for an answer to be given which would be adequate given modern naturalistic criteria for adequacy.

There are many answers which are adequate if the word is taken in the broad sense. For example, Thomas Aquinas (and many others) claimed that there would be no value-ordered world *at all* if there were not a God. This constitutes a direct answer to Bambrough's question. If God did not exist then, according to theists, almost nothing would be the same as it is in the present world. A world with no *meaning* would exist if God did not exist.

There are other answers as well which could be given. If there were a world in which every good act were immediately followed by punishment of the good-doer and the cancellation of the good that was done, no one would suppose that the world had been created by God. Of if there were intelligent and good-wishing beings but those beings had no power whatsoever over the things that went on in their bodies, so that when they wished to hug their neighbors they strangled them instead, no one would think that that world was created by God. Or if the world had intelligent beings but was a hedonistic paradise in which wishes for sensual pleasure were followed immediately by gratification like the world the fox and cat persuaded Pinocchio to seek, no one would suppose that there was a God. In all these cases a designer would be required, but not God.

It does not require much imagination to give other examples of what worlds not created by God might be like, hence it must be that Bambrough and Flew and other philosophers who ask this

² Renford Bambrough, *Reason, Truth and God* (London, 1969), p. 52.

question have something special in mind. If what they have in mind is "In what respects would a world created by God be different from this one?" the theist response has to be that such a world *might not in any basic respects be different from this one*. The present world is a very good illustration of what a world created by God would look like.

The Bambrough question is not neutral at all, it is a "trap" test like the requirement that theological propositions be verifiable in terms of sense experience or that language in theistic propositions be used in the same literal sense as that used in speaking of finite things. To answer the question in the terms proposed is to renounce the religious belief-system itself.

Thus even if religious believers win the argument that their belief-systems are empirically meaningful, there is no hope that they will win the battle for truth if naturalistically biased truth criteria are used as the test to distinguish truth from falsehood. This leaves only three options for the religious believer, to try to accommodate religion to a naturalistic world-view, to give up the quest for rational justification for religious belief but continue to believe, or to challenge the claim of naturalism that its view of reality and its methodology for evaluating truth claims are adequate to human experience. The first two options will be taken up in order.

Occasionally one encounters the view that the verification criterion has been liberating to religion, even that it has helped religion to return to its proper sphere, that of eliciting and strengthening value commitments rather than stating facts. According to John Herman Randall, religion offers no descriptions of fact, no truths about this world or some other world not attainable by the standard methods of science. Its function is to offer a set of symbols which will help us "see" something about our world. "A religious symbol unifies and sums up and brings to a focus men's long and intimate experience of their universe and of what it offers to human life."³

The world made manifest by religion is not an additional realm of fact but results from a particular way of seeing a world entirely disclosed in its factual aspects by science. Religious genius, like poetic and artistic genius, opens up for us things we have not seen before, possibilities for using our

eyes and ears and lives that we have not noticed before. Thus Randall in no way means to denigrate religion, but to remove it from the competition with science as a source of factual knowledge and to liberate it for its proper task in the values realm.

The proposal is a radically revisionist theory of the nature of religious-belief systems. While it preserves an essential feature of religion, its concern with providing a path of life for believers to follow, a way of salvation, it denies both the existence of supernatural beings (and aspects to human beings) and any divine influence on the conduct of nature and human affairs. Clearly this theory does not *describe* how typical religious believers themselves look upon their belief-systems but is a proposal that religious believers from now on look at their belief systems in this way. This proposed revision of the nature of religious assertions has not been widely accepted by religious believers, as might be expected. Clearly most religious believers regard the claim that their statements about God provide information about the world and man as essential to their belief-system. Religious matters are not, for them, merely matters of ultimate concern but also of theory of reality.

The second option is to consider whether it might be possible to claim that religion provides an autonomous source of knowledge. It might be defensible to claim that there is a methodology and way of describing that is appropriate to scientific truth and a methodology and way of describing that is appropriate to religious truth, and that each kind of truth has a method of justification internal to the truth-system. You should do science parsimoniously, but it might be all right to be a bit more generous when you do religion. Neither type of activity should dictate to the other how truth questions should be answered.

There are two major forms of this approach, that suggested and later abandoned by Alasdair MacIntyre, which apparently offered a frankly supernaturalistic and neo-orthodox version of religion, and that offered by D. Z. Phillips, which entails drastic modifications in what religious believers have typically believed. It is admittedly difficult to get very clear as to what either was advocating since both agreed that the attempt to translate what they were saying when they spoke

³ John Herman Randall, Jr., *The Role of Knowledge in Western Religion* (New York, 1958), p. 117. Other examples are provided by Paul F. Schmidt, *Religious Knowledge* (Free Press, 1961) and R. B. Braithwaite, "An Empiricist's View of the Nature of Religious Belief," *The Eddington Lectures* (Cambridge, 1955).

religiously into ordinary secular language was to distort their meaning.

The movement which Kai Nielsen has styled "Wittgensteinian Fideism"⁴ arose from a combination of a philosophical movement associated with the later writings of Wittgenstein—typified by the claims that language is basically all right as it stands and that each of the language-games which are played have their own standards of intelligibility and rationality—and a theological movement associated with the writings of Søren Kierkegaard and Karl Barth, who shared the belief that revelation is the sole source of religious truth. The common ground is the emancipation of religious language from any connection with science or rationality, hence agreement on the autonomy of basic language systems. The role of philosophy is neutral, to analyze and clarify the basic language-systems to keep them from interfering with each other, but not to bring an orderly combination out of them. Reconciliation is metaphysics, and metaphysics is a game which cannot be played. Religion, ethics, science are all autonomous and valid realms, to be protected from mutually interfering with one another.

Religiously this position is similar to that taken by the early Reformers except that the reformers merely separated knowledge of God which man has from nature from *saving knowledge* which he obtains by revelation, while the new breed of fideists claims that man acquires no knowledge of God whatsoever through sense experience and the orderly reflection thereon. In the realms of empirical fact man draws an absolute blank when he tries even to *think* of God, let alone when he tries to show that there is one. Both religion in its claims to revealed truth, and science in its claims that the empirical world scientifically understood has no room for the hypothesis that there is a God, are to be accepted.

The fideistic claim is a repudiation not only of the efforts of philosophers to ground religious claims by reason, but also of traditional religious appeals to miracles, divine actions and the like as evidence for the truth of religion. The ties were broken not only with the classical proofs for God's existence but with all attempts to use God as an explanatory hypothesis for nature and history as a whole or for particular events or chains of events occurring in nature or history. Religion appeals solely to religious authority, or at least will come

to do so when the theologians are led out of the fly-bottle by proper philosophical analysis.

In presenting Wittgensteinian fideism my goal is not to show that it is incoherent, for this is not in question—it *could* be played—but to show that it fails to account for how religious belief-systems actually work. Theology has not been and is not now practiced in the manner described.

Religious assertions, Alasdair MacIntyre claims, are not hypotheses. To say that they are hypotheses would be to clash with the way they traditionally are held by believers, he claims. The system of religious belief rests upon authority. "The only apologia for a religion is to describe its content in detail: and then either a man will find himself brought to say 'My Lord and my God' or he will not."⁵ It is "tis" or "taint." Elsewhere he says, "to ask for reasons or a justification of religious belief is not to have understood what religious belief is" (p. 208).

What is true of religion is true also of science and morals. "Of science and morals it can also be said that one can justify particular theories or prescriptions, but that one cannot justify science as a whole in non-scientific, or morals as a whole in non-moral terms. Every field is defined by reference to certain ultimate criteria" (p. 202). In religion that criterion is authority.

It is pointed out by MacIntyre that he does not mean to say that religion does not say things about the world, merely that what religion says cannot be justified by means of appeals to evidence or arguments drawn from the ordinary world. Of statements about God's love he says, "We do not offer evidence for these statements, we offer authority for them. We point to the state of the world as illustrative of doctrine, but never as evidence for it" (p. 201). "Our ground for saying it is that we have the authority of Jesus Christ for saying it: our ground for accepting what he says is what the apostles say about him; our ground for accepting the apostles? Here the argument ends or becomes circular; we either find an ultimate criterion of religious authority, or we refer to the content of what authority says" (p. 200).

But is it really the case that religious believers typically refer questions about religious belief to Jesus Christ as authority? Sometimes believers may say that they believe what they do because Jesus said so, but the statement bears inspection—do

⁴ Kai Nielsen, "Wittgensteinian Fideism," *Philosophy*, vol. 42 (1967), pp. 191-209.

⁵ Alasdair MacIntyre, "The Logical Status of Religious Beliefs," *Metaphysical Beliefs* (London, 1957), p. 205.

believers claim to trust Jesus as authority simply on the basis of his claims to embody authority or is it that they think that they have good reasons to think that his claims are valid? In other words, is belief in the authority of Jesus *itself* believed to be justified in terms of some other criterion? Official religion, at least, holds that belief in Jesus can itself be justified in terms of proofs and evidences. This is true of both Catholic and Protestant thought, except for fideists.

If MacIntyre's theory were true we should expect not only that religious believers should never attempt to justify their belief that Jesus Christ is the authority who should be accepted, but also that the content of religious belief would never be modified unless the words of the authority plainly support such a modification.

In point of fact, however, religious believers, including those believers who were historically closest to Jesus, do not justify the content of their beliefs just by appealing to authority but make attempts to prove that Jesus is divine, and, secondly, they have significantly modified their beliefs under the pressures of changing secular views in metaphysical age after age. The religious game is just not played in the way that MacIntyre leads us to expect.

In this connection it is instructive that while the New Testament Gospels preserve records of what Jesus said, almost nowhere in the remainder of the New Testament is he quoted as authority for belief. Apparently the apostles who wrote the various letters to Churches did not use Jesus' words as the basis for religious belief, but relied on the claim that certain events were signs of God's activity. When John the Baptist's disciples asked Jesus if he was the one to come, he told them to look at the signs—the lame walk, the deaf hear and so on. The early followers of Jesus told their hearers that God had raised Jesus from the dead and that *for this reason* they should believe in Jesus and be saved. The Catholic church has always claimed that its miraculous universality and its freedom from error on matters of faith and morals are evidence that it has supernatural backing. These rather typical religious claims are simply illustrations of the fact that religious believers have typically maintained that there is *evidence* which justifies accepting the belief-system. Moreover, they have offered philosophical arguments in support of the religious world-view.

The second thing to notice is that the content of religious belief is subject to changes that would be

unexpected on the fideistic theory. If MacIntyre were correct, religious assertions should never change unless new discoveries about Jesus' teachings are made.

Jesus apparently believed that Moses wrote the Pentateuch and presumably believed that history had the kind of time span consistent with the more recently proposed 4004 BC chronology, but religious believers do not generally accept such things as factual assertions even though Jesus apparently believed them. MacIntyre claims that historical events *can* be believed on religious authority. He says that "certain past events can be part of a religious belief, that is that they can be believed on authority" (p. 207) in speaking of the empty tomb. Because this is so, he says, any newly discovered document claiming to establish that the resurrection did not take place would be rejected as forgery. Why do not believers then accept *all* the things that Jesus said as factually true on the MacIntyre thesis? He gives no satisfactory account for how the believer selects which facts to believe, and he cannot do so without violating his fideistic thesis.

Religious claims and scientific claims seem to conflict at times, but the fideistic theory seems not to permit using an answer from the religious realm to settle a science question or vice versa. Does this mean that religions should make no statements about scientific fact questions? No, MacIntyre says that certain fact statements are to be believed on religious authority, and this is consistent with his fideistic thesis that science and religion have autonomy. Should the outcome of apparent conflict be that a person should accept conflicting answers, one from science and one from religion? Religious authority says that the earth is young, while science says it is old? Should we alternate days or should one answer be believed in church and the other in school? If neither science nor theology can criticize the other it seems logical to say that both should stand in their proper spheres.

The trouble with the MacIntyre thesis is that religious believers *do* try to bring coherence into their whole system of beliefs, religious, moral and scientific. They do not believe on religious authority that Moses wrote the Pentateuch, or that creation took place 6000 years ago, no matter what religious figure says these things, even if they believe that the person who said them is divine. They search for coherence as though they thought that there is an overarching system of thought which may begin with religious teachings but which must be

developed and tested in terms of its applicability to the whole of human experience. They believe that the games must not only be internally consistent but must be played consistently with each other, a belief no Wittgensteinian fideist can justify.

In elaboration of this point I shall discuss some passages written by D. Z. Phillips. While affirming also that each language is subject to its own criteria, Phillips is forced also to admit that religious language is subject to correction by other criteria, apparently. This seems to be the import of the following statements: "I have argued that religious reactions to various situations cannot be assessed according to some external criteria of adequacy. On the other hand, the connections between religious beliefs and such situations must not be fantastic. This in no way contradicts the earlier arguments, since whether the connections are fantastic is decided by criteria which are not in dispute."⁶ Also, "When what is said by religious believers does violate the facts or distort our apprehension of situations, no appeal to the fact that what is said is said in the name of religion can justify or excuse the violation or distortion" (pp. 98-99). He, like MacIntyre, believes that religious beliefs often concern the world, his claim is merely that they are not justified by appeals to the facts of the world.

It is natural to ask whether the foregoing passages maintain this distinction. If religion claims things which are fantastic, how can one reject these things on the grounds that they constitute a violation or distortion of the facts without by that very act of rejection showing that an alien criterion has been used to judge religious claims? Suppose, to cite an example from Phillips, that a boxer crosses himself before entering the ring because he thinks that this will protect him from harm. He is committing a blunder, according to Phillips, because he is violating well established views of causality. "We can say that the people involved are reasoning wrongly, meaning by this that they contradict what we already know. The activities are brought under a system where theory, repeatability, explanatory force, etc., are important features, and they are shown to be wanting, shown to be blunders" (p. 102). But for Phillips to say this is to impose an alien theory of causality. It is characteristic of some religious belief-systems to claim that in fact there is divine intervention in events. Moses expected God to do certain things

when he held out the rod, and Elijah expected fire from heaven to consume his sacrifice. To claim that God does things in nature and history is to make a religious claim, it is part of the religious way of viewing life.

The problem is that while Phillips in these passages is reporting a fact, that religious claims about divine causality and scientific claims which supply natural causes for these same events are conflicting explanations, his thesis about the independence of languages cannot account for this fact. Many religious believers have demythologized claims about divine activity because they do not believe that such events happen, they believe instead that events are caused by natural forces or by human agency because they have come to accept a scientific view of the universe instead of that part of the religious view they once held. But unless one wants to claim that the disappearance of belief in the miraculous is based on religious reasons (and an argument of this sort could be developed, however Phillips does not develop it as far as I can see), to reject the belief that crossing oneself protects one from harm is to subject religious beliefs to an alien criterion.

In short, Phillip's concession to the elimination of the fantastic is a partial abandonment of Wittgensteinian fideism. It looks rather more like the kind of theory which results when one practices, as I would, the once-honored activity of attempting to harmonize what we say about the world in our various languages by building comprehensive philosophical systems.

It has been my contention that two features of the way religious language develops supports this latter way of structuring the problem and deny the adequacy of Wittgensteinian fideism as a model for analyzing religious language. In the first place, religious believers have always presented what they take to be evidence in support of their views. They argue from extraordinary events, unusual persons, historical movements, patterns of design, and they offer explanatory hypotheses for religious experience and of moral demands. Secondly, religious views have varied relative to changes in scientific and metaphysical frameworks for interpreting ordinary events. For these two reasons it cannot be successfully maintained either that religions do not seek justification for their claims to truth or that religious beliefs are subject only to religious criteria for truth.

⁶ D. Z. Phillips, *Faith and Philosophical Inquiry* (New York, 1971), p. 98.

To put it differently, Phillips is right in maintaining the thesis that religious belief-systems cannot be justified by "alien" criteria. The truth of a religious belief-system is shown by religious experience and by no other type of appeal (assuming that a coherent metaphysics can be developed). This does not mean, however, that all the parts of the religious belief-system are justified only in terms of an appeal to religious experience or religious authority. Religious beliefs are justified also by the realities believers see, not just by the way they are told to see things by their religious community and in their religious language. They are heavily influenced by science in how they see the religious world, and the particular modifications in traditional belief that Phillips himself acknowledges are evidence of this.

Phillips rejects the efficacy of petitionary prayer, the belief that life goes on beyond death, the belief that God intervenes in the world, the idea that the last Judgment is an event, and the belief that all evil will be made right some day, in all these respects providing clear evidence of the intrusion of science upon his religion. Nielsen is right in noting that what Phillips advocates is "a radically reconstructed Christianity masquerading as a neutral conceptual analysis of Christian discourse."⁷ He is wrong, however, in finding anything inherently contradictory in fideism.

Nothing I have said or could say would refute the possibility of being a consistent fideist. What I have tried to show is that fideism is atypical of believers, and that it is not followed by either of the fideists selected for examination. Fideism *could* be

played, however. A fideist could say of every belief that he has, "this was revealed to me." He could reject scientific regularity claiming that God is the cause of every event, that every event has a particular reason and that it occurs to serve that particular purpose. He could say that the eight ball will go in the side pocket if God sends it there, regardless of the skill of the pool-shooter. This kind of outlook would result if every belief were derived from authority. It is plain that this is not the system expounded by either MacIntyre or Phillips, but it could be defended consistently.

The proper conclusion to draw is, I think, that if religious belief-systems are to be defended it must be in terms of the same old warfare between science and theology that has been waged through recent centuries. Religious belief systems must be defended through the development of a comprehensive metaphysical theory grounded in religious experience, which theory is not only adequate to all the known facts but is claimed to be superior to the view of reality based on acceptance of the verification principle that Nielsen supports. It may have taken fifty years to establish that the attempts to dissolve metaphysics associated with Wittgenstein (in both early and later stages) and with Karl Barth rest upon a mistaken account of how religious and other language systems function, but there are signs of growing recognition that the battle has returned to the fields of yore. It is somewhat comforting to know that those philosophers who have argued the pros and cons of religious belief were, after all, really *arguing* about something after all

University of Delaware

Received April 4, 1975

⁷ Kai Nielsen, *Scepticism* (New York, 1973), p. 32.

VI. ON PRACTICE

JOSÉ FERRATER-MORA

IT is not unreasonable to assume that a theory and, in general, any cognitive undertaking does not shape up and develop entirely by itself, in complete independence of strictly non-cognitive contexts, such as systems of beliefs, ideologies, world views, and even the conditions—historical, social, economic, national, etc.—out of which the contexts arise. In this sense, theories are heteronomous.¹ Yet they do not seem to be completely heteronomous in so far as a certain degree of what can be called “internal evolution” can be detected in the formation and development of theories. The degree of this internal evolution, which makes a theory at least relatively autonomous, varies according to the type of theory considered. Thus, in the formal sciences—probably because they are not *stricto sensu* cognitive—the degree of autonomy is maximum, and, as some would say, complete, although contexts may occur, as with the natural sciences, that often mold the ways in which the formal sciences develop. There is less autonomy in the natural sciences than in the formal sciences. In the social sciences and, for somewhat different reasons, in philosophy, the degree of autonomy is considerably reduced, and, according to some authors, practically nonexistent. In any case, however, one can admit that there is some autonomy, and in particular, what we may call “formative autonomy.”

It may, and often has been argued, that heteronomy only concerns the formation and development of theories, namely, the concrete ways in which knowledge is produced; heteronomy therefore, pertains to what is usually called “the context of discovery.” Such a context can oscillate between a “minimal context”—conceptual frameworks, assumptions not too far removed from, or too alien to, the theories—and a “maximum context.” The

latter does not necessarily include “everything,” for otherwise interminable descriptions of any “process of discovery” would ensue. It may include, however, such contexts as ideologies, world views, etc., on the one hand, and political and social institutions, economic structures, etc., on the other hand. It is then contended that no matter how much these contexts may clarify the discovery of a theory, they will say little, or nothing, about whether or not the theory—or the thesis, or the hypothesis—at stake is acceptable, or valid, etc. The latter point seems to be within the jurisdiction of the so-called “context of justification” or, as I will also call it, “context of validation,” where factors contributing to the formation and development of a theory or, in general, of knowledge are not normally taken into account unless they are a part of the methodological procedures. Accordingly, it is affirmed that even if knowledge were formatively heteronomous, it would still be autonomous “validatively.”

Neat distinctions are like well focused pictures: their degree of “definition” is high. Hence they are eminently persuasive, for we can always appeal to them when the images begin to get blurred. The question is whether the price paid for a neat picture is too high.

At the present stage, it does not seem to be too high. The distinction between the context of discovery and the context of justification (or validation) serves to warn us against perfectly unnecessary confusions, worth no price at all. Furthermore, as Wesley C. Salmon has emphasized, it is possible to maintain the distinction between the two contexts without assuming that they exclude each other. The same “factor” can thus be pertinent both psychologically—and, I assume, also sociologically—and logically to a given theory or

¹ “Extratheoretical contexts” may impinge on theories in many different ways. To begin with, a distinction can be made between formation and development, which are diachronical, and structure, which is synchronical. Furthermore, whereas formation and development can be accounted for genetically, such is not necessarily the case with structure. Secondly, *X* may impinge upon *T* in various ways: it can be the (direct, indirect, concomitant) cause of *T* or the frame within which we are supposed to understand *T*. In the latter case we may call *X* a “context”, *C*. Still, *C* can be a context of *T* in so far as the conceptual framework of *C* includes the conceptual framework of *T*, or in so far as appeal to *C*, or to parts of *C*, may allow us to understand *T*, or parts of *T*, etc. Needless to say, notions such as “conceptual framework”, “understand”, are in need of clarification. I hope that the “context” within which these pages are written is reasonably clear.

hypothesis. A factor common to both contexts does not obliterate the distinction proposed since it functions within each context in its own way. If such is the case, the price to be paid for the distinction between the context of discovery and the context of justification is quite a bargain.

Furthermore, if this were not enough, the distinction in question avoids the pitfalls of piling everything in a kind of conceptual heap. Such heaps emerge as soon as we try to relate a theory to rather "distant" contexts, as is the case with social and political ideologies, not to mention the concrete conditions from which these ideologies are supposed to emanate, whether reflectively or reactively. Let us call any theory '*T*', and any context of the kind alluded to '*C*'. It is only reasonable to contend that even if *C* were the cause of *T*, *C* could not explain *T* other than genetically, and *C* would not logically imply *T*. *T* is not derivable, in any sense of the word, from *C*, and in this respect *T* is autonomous with respect to *C*. If *T* could be derived from *C*, then it could be shown that it is so derivable. If, for example, it is held that a political ideology, *I*, forms the context for the justification, or validation, of *T*, or of the statements in *T*, then these statements should be derived from *I*. If they are so derived, then the justification or validation, and not only the real production, of *T* by *I* will be in perfect order.

To the best of my knowledge, at present, those derivations which have been made were either unsatisfactory, or they have failed to distinguish between the contents of *T* and the contents of *I*—sometimes *I* is even a rationalization or a specification of *I*. Even in such a case, however, it is doubtful that non-genetical derivations—to say nothing of logical ones—have been performed. We may trace in *T* a number of evaluations, prescriptions or orientations which originated in *I*, and we still wonder to what extent any of them helps to validate *T*.

Now, the distinction between the context of discovery and the context of justification works adequately only within the rather restricted conditions for which it was devised, namely, conditions in which derivability, as a tool for logical and methodological reconstruction, is of paramount importance. This, however, presupposes a certain notion of "theory," which is by no means mandatory. Other theories concerning "theories," and, in general, concerning "knowledge" may enter this competition in the two areas which I am now hurriedly exploring: the philo-

sophy of science (as well as the philosophy of philosophy) and the sociology of knowledge. One of these "non-orthodox" theories I am quite willing to defend.

My defense ought not to be understood as a suggestion that one go to the other extreme, like those who have rejected the distinction between the context of discovery and the context of validation. Such a rejection was possible only because one of the contexts was eliminated and the latter, in turn, happened when attention paid to the relations between the propositional structure of a science and its extrascientific contexts led to the conclusion that this structure is theoretically justifiable—and not only genetically describable or explainable—by dint of certain empirical conditions, particularly psychological, social and "institutional." Thus, justificationist absolutism has been replaced by psychologistic or sociologistic absolutism, and I sometimes wonder which one is the better, or the worse, of the two.

My plans are easier to announce than to fulfil, but I will try against all odds.

Let us consider, to begin with, what type of relations other than derivability or causal production, can be established between a theory, *T*, which tries to be "objective," and thus "autonomously" justifiable, and a context, *C*. Since not all theories, and certainly not all contexts are of the same kind, it is likely that even if a new given relation is discovered, it will function quite differently according to the kinds of *T*'s and *C*'s considered. I shall leave these annoying details to non-philosophers.

It is sometimes held that *C* is a part of the validation, or justification, of *T* in so far as it is a necessary, but not a sufficient, condition of *T*. This view seems acceptable to me, but trivial. If, for example, a social ideology is the necessary, but not the sufficient, condition for a scientific theory, then the ideology will be part of the discovery, formation and development of the theory, but it will neither help nor hinder the validation of the theory.

My plan is a strategy, and it is carried on in two tactical steps, which could be called "maneuvers." The first maneuver consists in starting with what I have called "theory," in order to try to relate it, intelligibly, or rationally, to as many contexts as possible. The second maneuver starts with some context in order to see how theories can be brought to bear on it.

First maneuver. A viewpoint on "theory" somewhat

different from, although not necessarily incompatible with, the one ordinarily associated with the distinction between the context of discovery and the context of justification (or, again, validation) can try to trace in the theory some elements which, to begin with, may be not too far removed from the theoretical structure. Among these elements are expectations of results which are not logical derivations of the theory—as, for example, the capacity to adapt the theory to situations which had not been previously considered; the possibility of modifying the theory changing (generally, enlarging, but sometimes restricting) the meanings of some of its terms, or the scope of some of its concepts; the introduction, or substantial modification, of an observational language, or parts thereof. These “elements” are a function of contexts which operate as theoretical assumptions and which do not need to be removed from the theory as if they were “only” empirical conditions of its production. The distinction (between the context of discovery and the context of justification) thesis does not eliminate altogether the afore mentioned elements, but tends to mistrust them. To be sure, it tends to consider an observational language as extra-theoretical. I am aware of the fact that it is not always easy to pin down the “elements” or “factors” just alluded to, and that when they can be pinned down, it still remains difficult to ascertain how they are related to the theory. The innumerable debates on the nature and function of the so-called “assumptions” or “pre-conceptions” show that this matter is far from being clear in the mind of everyone (they are not clear in my mind). Nevertheless, if there are some such assumptions or pre-conceptions, they function theoretically or, in any case, cognitively. When the latter are considered, a theory becomes richer, namely, it increases its cognitive content. Consider in this respect two quite different theories: the theory (principle, or law) of inertia, in its Galilean version, and the nominalistic theory (or doctrine) of Occam. The principle, or law, of inertia, as it is ordinarily formulated in basic Physics textbooks, does not seem to contain assumptions or pre-conceptions of any kind, unless we count among them stipulations of the “in ideal conditions of absence of friction . . .” type. It has been shown however, (by Hanson and Koyré among others), that the conceptual content of the principle is richer than its textbook formulation. Similarly, Occam’s nominalistic doctrine has theological pre-conceptions and yet this doctrine is not

derived from them, since the same pre-conceptions could serve to outline a non-nominalistic theory. These pre-conceptions, on the other hand, are not indifferent to the arguments pro and con Occam’s views. In point of fact, without these pre-conceptions, Occam’s doctrine would cease to be what it is, and the arguments adduced in its defense would probably change drastically.

It should be pointed out that the assumptions of a theory can become oppressive rather than liberating—or impoverishing rather than enriching. An increase in the cognitive content of a theory is not obtained by eliminating all assumptions, but rather by trying to find some others. It often happens that a change in the theory increases its cognitive content.

Other elements or factors which it is reasonable to assume are a part of the structure of a theory contribute to its validation. Such elements include: the community of researchers; the consensus by virtue of which a theory is accepted, or rejected; and the possibility of communicating the content (meaning) of the theory and of its terms.

I do not understand by “the community of researchers” a given group of human beings doing research. Although there must be such a group, concretely operating within given social and cultural conditions, what matters now are rather the rules which make such research acceptable or subject to intelligible discussion by the members of the group. These rules are not only methodological rules, but also, and perhaps above all, rules which circumscribe the content and scope of the research. Fundamentally, they are the series of steps and “maneuvers,” both positive and negative, which Lakatos has described as “scientific research programmes.” To some extent the methodology of such programmes coincides with what I have previously called “assumptions” or “pre-conceptions,” but the latter has a stronger pragmatic character than the former. This does not mean, however, that they must be rejected as not being a part of the ensuing theories. Pragmatic reasons include important elements concerning the context of discovery. Such reasons are coincident rather than incompatible with validation. They might also be the locus where the two contexts—discovery and validation—meet.

Something similar can also be said of the two other “elements” or “factors” already mentioned, particularly the element of “consensus,” a rather ambiguous and even equivocal term since it includes the notion of disagreement. Unless we

admit a quasi-Platonic criterion of ideal objectivity which allows us to decide whether or not a given theory is acceptable or the arguments adduced in its favor are plausible, or whether a given observational language is adequate or inadequate; or unless we subscribe to some notion of "irresistible evidence," we cannot help but conclude that what is called "objectivity" is, in the last resort, the same as what is inter-subjectively agreed to. This does not mean, of course, that a consensus is reached by vote counting, or by tyrannical rule, or by more or less "authoritarian" and "authoritative" influences—even if this is just what sometimes happens. No matter how much one argues, reasons, computes, or appeals to empirical data formulated in an observational language—which, by the way, may itself be the subject of some "consensus"—what in the last resort makes a proposition acceptable or not acceptable is the consensus, or agreement, that it is, indeed, acceptable or not. This consensus functions within the rules laid down, implicitly or explicitly, by the community of researchers by virtue of habits engendered by a multitude of common experiences. The consensus is tied up to the possibilities of communication and, therefore, to some "linguistic" context. The latter must be understood in a rather wide sense, as covering all the possible criteria of rationality. Thus, the "linguistic context" which is associated with "elements" still wider in scope than those mentioned up to the present, will stretch the context of validation to its limits—something I am sure some will consider frankly intolerable.

We must now consider what has been named "world views," an expression whose meaning is so nebulous as to suggest that we might be wise in dispensing with it. I believe, however, that we can utilize it by considering it as a basic and very large frame of beliefs and evaluations. On the one hand, a world view provides orientations for the understanding of, and action on the world (generally, selecting some assumedly fundamental features or elements), and, on the other hand, and concomitantly, it helps us to perform intellectual, or practical, decisions when diverse and perhaps conflicting interpretations or norms are presented.

Now, although both function in similar ways, a world view must not be confused with an ideology. The latter is usually confined to beliefs and evaluations concerning the political and social environment. For "preferential reasons" I will concentrate on the so-called "ideologies," regardless of whether they are of a "conformist" or of a

"nonconformist" type; namely, whether they are, in Mannheim's vocabulary, "ideologies," reflecting a social system more or less faithfully, or "utopias," reacting to a social situation by proposing a new ideology. In both cases, ideologies are linked to social conditions, without which they would have little, or no, sense. For this reason, it is sometimes said that if a theory is socially conditioned, and if this condition is a part of the context of validation, then the theory will have to be validated within the scope of such conditions.

Now, I believe that this has to be taken cautiously. Not even those who most staunchly defend theories as a function of social conditions would, in their sanest moments, dare to say that these very conditions validate, or fail to validate, the theory. This is true even if we take "validate" in an ample sense, rather than equating it with "logically derive." For example, an appeal to social conditions cannot validate or invalidate an optical theory even if the latter has emerged within given social conditions—for the latter could give origin to several optical theories. Thus the notion of "social condition" really expresses a condition rather than a cause.

It is quite another question to ask whether an "ideology" can, or cannot play a part in the procedures which validate a theory. I am thereby raising the problem of whether there is any relation, not strictly genetical in character, between what we may call "ideological conceptualization" and "theoretical conceptualization" proper.

It can be claimed that an answer to the above question depends upon what kind of theory we have in mind, since, as I already pointed out, it is customary, and indeed correct, to maintain that there is a closer relation between ideologies and social theories than there is between ideologies and theories in the natural sciences, where the relation becomes quite problematic. Now, unless we wish to admit that there is an essential rather than a gradual difference between various types of sciences, it will be mandatory to answer the question unequivocally, so that what is asserted applies to all kinds of theories, no matter how different they are among themselves.

My opinion is that the notion of justification, or validation, can be expanded from a strictly methodological and "reconstructionist" area to a much wider area, where validation does not follow the theory once it is constituted, but rather actual constitution of the theory and its validation overlap. Thus, classical preformist theories in genetics

adduced facts which validated the theory, but at the same time the possibilities of validation—including, among other things, the selection of an observational language—were not alien to the ways in which the theory was formulated. In present day genetics a theory has been developed according to which the nucleus of each cell in an organism possesses all the necessary information to reproduce the entire organism. If many cells prove incapable of such reproduction, it is only because the nuclei are programmed for the performance of specialized functions. The theory alluded to is largely based upon research on the composition of chemical substances which transmit, and select, information through molecules of ribo-nucleic acid. Thus we can say that the results of this research not only serve to confirm the theory, but at the same time condition its structure as well as its development.

It may be claimed that the orthodox notion of validation, understood as the logical and methodological reconstruction of a theory, takes into account confirmability, as well as its degrees. But what it does *not* take into account is the fact that the very structure of the theory—not only its actual development as a process of “discovery”—is molded by the possibilities of validation. Above all, it does *not* take into account that validation can also include a research programme which serves as a framework for the formulation of given types of theories. The steps taken to constitute and develop the programme are parallel to the phases of theoretical construction, and since these steps include procedures of validation, without which they would be mere gropings in the dark, it is reasonable to conclude that at least part of the so-called “validation” is not alien to the steps taken to constitute and develop the programme. Therefore, validation cannot be viewed as a mere logical and methodological reconstruction.

Some readers may say that the notion of “validation” is so tenuous that it scarcely deserves the name any longer. I agree, but this was precisely one of the aims in my strategic-tactical maneuver: I am trying to suggest that such a notion can be expanded to the extent that when it reaches certain limits we are no longer certain of whether we are dealing with validation or with discovery. They may be two sides of the same coin. In order to lessen the intellectual shock which this invitation to conceptual indeterminacy might produce, semantic stipulation can be laid down which will appease hard-boiled methodologists. I suspect,

however, that it would be best to consider particular cases and ascertain whether, and to what extent, the notion of validation can be expanded without losing its meaning.

Now, to what extent can an ideology be part of the validation of a theory? I will offer some suggestions which I hope are not entirely foolish.

An ideology can be considered as an expression of “human practice,” and specifically of what Jürgen Habermas has called “interests,” defined as basic orientations rooted in the conditions that make reproduction and self-constitution of the human species possible. In any case, practice provides certain orientations and establishes certain aims for the development of scientific research programmes and, in general, of knowledge. These orientations and aims contribute to the validation of the theories developed in the program in so far as they single out, or select, the elements to which procedures of validation apply. In *this* sense—but in this sense only—it can be said that “practice” orients theory, and plays a part in establishing the criteria for the validation of the theory.

Once this is acknowledged, however, it must be pointed out that the procedures used in validating theories cannot be derived even indirectly, from “practice” and from ideologies. Firstly, as I surmised a while ago, parts of theories develop “autonomously,” and the more they do so, the more they are validated “autonomously,” that is to say, “theoretically,” or “scientifically,” and non-“ideologically.” Secondly, even in areas where ideological orientation is strongest, as in social theories, and possibly in some philosophical theories, the criteria of validation, albeit more ample and flexible than those proposed by “traditional” and “orthodox” methodologists, cannot be extended indefinitely. Within the frame of an ideology a certain number of criteria, such as explanation, prediction, intelligibility, richness of content, practical applicability, etc.; can be laid down, and in this sense the ideology contributes to what we may understand as “validation.” The ways in which these criteria function are “ideological,” rather than “pragmatical,” in so far as they are steps, no matter how tentative and rectifiable, which can always be justified within the framework of fulfilling the aims pursued. Although it might be claimed that the idea of a “pragmatic function” is, at bottom, “ideological,” it is so only to the extent that it becomes a criterion which may or may not work according to whether, and how, it is applied.

Second maneuver. Let us now consider the question from the other side.

I think that we can start from the following situation: theories and, in general, knowledge—but also systems of norms—are products of activities performed by human beings who live in communities and societies, and are conditioned by physical and biological factors—including the activity of adapting to, transforming, or “respecting” the world—and by economic relations of production, social structures, political institutions, national, tribal and racial forms of life, techniques, rites, customs, etc. All of these are “realities”—or “real situations”—as well as results of reactions to realities, and reactions to these reactions. Within these “realities” we find systems of norms and forms of life, ideologies, world views, etc., which, as pointed out earlier, reflect the aforementioned realities or express a state of rebellion against them.

It can easily be seen how difficult it is to distinguish between what “reality” is and what “activity” is; some of the factors, elements or conditions just mentioned fall under both categories. It is common to call “realities” a set of given situations, and to call “activities” what men do, or plan to do, in order to confront these situations, including the decision not to confront them at all. It is more confusing to make a distinction of this sort than not to make one at all; in some sense, everything I have been talking about are “activities,” even if they are called, redundantly, “real activities.” Among them are theories of all kinds, both scientific and philosophical, and, in general, cognitive acts, or “knowledge.”

From *this* point of view, both knowledge and theory can be characterized as an “activity.” If we summarize all the activities alluded to earlier under the ambiguous but handy name of “practice,” we shall be able to assert that theory, as a theoretical activity, is practice. In any case, knowledge is not a reflection of a transcendent, nor even of a transcendental world; it is not what may ensue from a purely “contemplative” and supposedly disinterested life, nor is it alien to the concrete conflicts and problems which human beings confront throughout their existence in this one and only world. There is no need for asserting that there is an impassable gap between theory and practice; indeed, if attention is paid to what may loosely be called “propulsing force,” the “primacy of practice,” specifically of “social practice,” seems almost inevitable. Thus we must acknowledge the paramount importance of ideologies in so far as they

express, conservatively, revolutionarily, or sceptically, the conditions of practice.

Does all this mean that nothing can any longer be called “theory”? Such a claim would go too far, for we might then take refuge in the “night where all cats are black” fallacy. The idea of a close relation between theory and practice emphasizes only that there are not two completely separated human worlds: the “contemplative” and the “active.” If we persist in using the expression “two-worlds”, we had better extend the two worlds over some kind of continuum; one of whose extremes we call “theory”—or knowledge, or science, or whatever—and the other “practice.” The suggested primacy of practice within this continuum indicates only that it is the locus where cognitive activities (among others) take place. None of this means that in order to validate a theory we must directly appeal to practice, or to some ideology, whether conformist or revolutionary, developed within the context of practice. We should rather appeal to a series of techniques—logical, epistemological, and, in general, rational—which have probably originated in practice, or even within the context of ideologies emerging from practice, but which have developed throughout history in a process of trial and error. Some of these techniques may, indeed, be chosen “for ideological reasons,” and in this sense their adoption may be considered as a step towards the validation of theories, but the adoption of a technique of validation is not to be confused with validation procedures.

From this viewpoint, therefore, the validation of theories is not completely independent from “practice,” but this does not mean that practice, or the “ensuing” ideologies, mechanically produce any techniques of validation. Furthermore, once the practical or theoretico-practical ends are established, the use of these techniques develop, or tend to develop, autonomously. Thus, one can choose a logical system in order to perform inferences, but the inferences performed are logical, not “practical,” and still less “ideological.” Because “interests” emerge within the context of practice and are often made explicit by ideology, one can, for example, postulate the need to formulate predictive statements. Prediction expresses then an “ideological preference.” Yet, the rules of prediction are not ideological—they are rationally established rules. If the need to formulate predictive statements is denied, then no set of predictive rules serves any purpose. Thus, one may conclude that validation is not equivalent to

prediction. But once prediction is rejected, and some other way of validation is chosen—as happened, for instance, with “understanding,” in the sense of *Verstehen*—it will be necessary to find out what kind of rules of “understanding,” as *Verstehen*, function adequately. The rules themselves will not be ideological—no matter how “ideologically” one chooses a method of “understanding” as opposed to a method of “prediction.”

Thanks to these admittedly subtle distinctions, it may then be possible—without subscribing to any set of eternal truths, or eternal rules—to maintain the independence of research programmes against any oppressive ideologies. After all, ideologies should submit themselves to criticism. Once their function as (positive or negative) rationalizations of practice is acknowledged, the question may be raised whether or not there are “preferential reasons” which make certain ideologies more appealing than others. Among the least appealing ideologies are those which are deceptively, or perhaps wickedly, claimed not to be ideologies at all. In ideological matters, however, the best policy is still honesty.

To be sure, anything that is said about practice,

including the statement that practice has a primacy over theory, is part and parcel of some theory. A kind of circle, not dissimilar to the so-called “hermeneutic circle” seems here inevitable. Now, although I do not believe that this circle can be eliminated altogether its disturbing character can be made somewhat less annoying by noticing that the theory in question is not, strictly speaking, a scientific theory, but an “ideological theory,” and one that honestly acknowledges its ideological and, as it were, “preferential” character.

Philosophy has been on the verge of dying a number of times, but each time she has surprised everyone, including those philosophers who declared her defunct, by producing yet another “tentacle” in her octopus-like structure. One of the tentacles of philosophy stretches out towards science to the point that there is a certain rather welcome continuity between philosophical theory and scientific theory. Perhaps another tentacle stretches towards ideology. Whether ideology and science can get together, at least to discuss their respective aims, through philosophy, is an open question which is philosophically worthwhile to pursue further.

Bryn Mawr College

Received January 7, 1975

VII. THE NON-SENSUOUS EPISTEMIC ACCOUNT OF PERCEPTION

J. B. MAUND

ACCORDING to one important philosophical tradition, an essential component of having a perceptual experience is the sensing of some kind of sensory item or sense-impression (variously interpreted as sensing a sensory particular, a *sensum*, *sensibillum*, *sense-datum* or as the having of a sensation sensory state or mode of sensing). Let us term such a view "the sense-impression account."

There is however an alternative account of perceptual experience, "the epistemic account," whereby perceptual experience is nothing but the acquiring of appropriate beliefs about the object perceived, or nothing but the thinking that one is perceiving the appropriate object. Macbeth is said not to be sensing dagger-like sense-data but rather to acquire the (illusory) belief that he was seeing a dagger. "He thought he saw a dagger but of course he didn't see one." Similarly the drunk who sees "double" does not have two sets of sense-data, e.g., two sense-data corresponding to two candles, but instead thinks that he sees two candles.

One of the few to carry out a detailed exposition¹ of the epistemic account is D. M. Armstrong. (To be accurate one should distinguish between sensuous and non-sensuous forms of epistemic accounts. My concern in this paper will be with the non-sensuous epistemic account although in referring to this account I shall use the briefer title "the epistemic account.") According to Armstrong perception is the acquiring of true or false beliefs about states of one's environment including one's own body. Veridical perception is the acquiring of true beliefs and illusory and hallucinatory perception the acquiring of false beliefs. Armstrong qualifies his account however for there are illusions and hallucinations where one recognizes the situation for what it is and one is not taken in. In these situations the perceiver does not acquire false beliefs at all but instead is said to acquire potential beliefs, i.e., an event occurs in the percipient which, but for the inhibitory

nature of other beliefs, would be the acquiring of a false belief. In this manner Armstrong claims to provide an account of those difficult situations of perceptual experience without belief, or at least without the right sort of belief. (G. Pitcher, who advocates another variant² of the epistemic account, uses a notion similar to "potential belief," that of "suppressed-inclination-to-believe.")

The difficulty posed for the epistemic account arises in those cases in which a person *Q* has a perceptual experience, e.g., of something's looking red to *Q*, but *Q* does not acquire any corresponding belief (or inclination-to-belief). In unusual lighting a ball may look red to me, even though I believe, indeed know, that it is white.

To overcome this difficulty it is claimed that, while in such cases the person does not acquire a perceptual belief that there is something red at place *u*, *Q* does acquire what is called a potential belief, or a suppressed inclination to believe that there is something red at place *u*. It is then claimed that an account of perceptual experience does not need to postulate sense-impressions for such cases but can generalize the account of acquiring belief to include the acquiring of a potential belief or suppressed inclination to belief.

But even if it is true that in such cases of experience-without-belief, a person always has a potential belief or suppressed inclination, I should argue that an account of what such potential beliefs or suppressed inclinations are, applied to the specific case of perceptual experience, would depend on the prior admission of sense-impressions. But in any case I think it can be shown that even if every perceptual experience does involve a perceptual belief (actual or potential), there are at least *some* features of some (and indeed very many) perceptual experiences which do not correspond to any appropriate perceptual belief, actual or potential. And nothing short of providing an account of *all* features of perceptual experience

¹ D. M. Armstrong, *Perception and the Physical World* (Boston, 1961); *A Materialist Theory of the Mind* (Boston, 1968).

² George Pitcher, *A Theory of Perception* (Princeton, 1971).

in terms of beliefs acquired, will do if the epistemic account is to be saved.

I

Very roughly the strategy of the upholders of the epistemic account is to present the difference between the epistemic account and sense-impression account in the following way: whereas the sense-impressionist will ascribe to a percipient's sense-impressions on a given occasion a range of features F_1, G_1, H_1 , etc., i.e., person Q has a sense-impression of something red, triangular and moving in front of him, the epistemic theorist proposes that Q acquires a belief *that* there is something red, triangular and moving in front of him. Whereas the sense-impressionist distinguishes between

- (a) having an impression *of* something red and triangular in front of him
and
- (b) being under the impression *that* there is something red and triangular in front of him,

the defender of the epistemic account rejects this distinction, claiming that (a) comes down to something very like (b). This way of putting the difference suggests that on the epistemic account

- (1) Q has a sense-impression that has features F_1 and G_1 cashes out as
- (2) Q acquires a belief that something is F_1 and G_1 .

My approach will be to argue that there are perceptual situations in which the alleged translation won't work; where we have a good reason to suppose that a sense-impression has at least *some* features H_1 that doesn't carry over into any appropriate sentence about beliefs acquired by the perceiver. And, to state it briefly, the proof of this rests on the claim that a given sense-impression with a range of features has *parts* each of which of course has a range of features, and the relation between part and whole of a sense-impression is such that it is not reflected (as it ought to be if the epistemic account were valid) in the relation between part and whole of the perceptual beliefs acquired. And the reason for this is that in general a perceptual belief acquired in perception is a belief that p , i.e., a belief that has propositional content (given by p); and the relevant part of a belief that p , must be some other belief (that q) whose propositional content q is part of, or is con-

tained in, the former propositional content p . In other words p must be a conjunction of q and some other proposition. For example, suppose Q has a sense-impression of two stars twinkling simultaneously (perhaps the cloud cover is broken only momentarily in different regions of the sky). The propositional content p of the corresponding perceptual belief acquired will be something like:

p = there are two stars there (place u) and there (place v) twinkling now.

Now the sense-impression involved may be said to contain parts, i.e., to be of a star twinkling at place u and of a star twinkling at place v . But its containing parts such as these can be interpreted on the epistemic account as its involving the acquisition of a complex composite propositional belief p whose parts are other propositional beliefs, e.g.,

- q_1 : there is a star twinkling now (t_1) at place u ,
and
- q_2 : there is a star twinkling now (t_1) at place v .

And it might just be plausible to interpret p as the logical sum of q_1 and q_2 . But however plausible it is to so interpret p in this example, there are other cases in which it is most implausible. The situations in which it appears to break down most dramatically are those cases in which a perceptual experience, say a visual experience, is a composite of a veridical and hallucinatory experience.

To illustrate, take the case of the twinkling stars. There are two important features of such visual experiences of stars. First of all, they contain illusory elements since the events one sees appear to occur "now" (i.e., at the time of having the visual experience) whereas they are mostly past events. Secondly, when one sees stars one can often see a number of stars as part of a definite pattern of stars, e.g., as a cross, a circle or (more imaginatively) as a centaur. And often at least such patterns are instant "gestalt" figures—one sees a circle or triangle of stars.

It seems clearly possible to imagine the following situation. A perceiver sees in a part of the sky a figure '8' composed of star-like sources of light. In fact it turns out that most of the sources are physical stars (and the perception of them is partly illusory, i.e., with respect to the time element) while one is a hallucinatory star (maybe it is a certain sort of after-image, maybe it is drug-induced) and yet another is a veridically perceived source of light—the event seen which appears to

occur now does in fact occur now (e.g., it is a man-made satellite circling the earth).

Now it seems to me that to handle such cases is a problem for any epistemic account that proposes to present a typical hallucinatory experience of Q as something like

" Q 's thinking that (believing that etc.) such and such is before him"

where there is in fact nothing there at all, and a typical illusory experience of Q , as

" Q 's thinking that something (which in fact is F) is G "

where G is incompatible with F .

Now the visual experience of seeing a figure-8 pattern of stars, the "whole" experience, is a composite experience with parts. Part of the experience is veridical—there are sources of light corresponding to most of the "seen stars"—part of the experience is non-veridical, hallucinatory let us say. If we gave an epistemic analysis of the experience of seeing stars in a figure-8 pattern, we must say that the perceiver is acquiring a belief, a false belief, that in a certain region of the sky, that there is a number of stars forming a figure-8 pattern.

In actual fact, let us suppose, there is an S-shaped pattern of stars in the region of the sky being considered (most of the rest of the sky is behind a cloud cover let us also assume). Let us also put aside the question of time-gap (we could equally as well have chosen another example in which we had an S-shaped pattern of objects close at hand, e.g., an S-shaped pattern of light bulbs on a tree). The percipient however, sees not an S-shaped pattern but an 8-shaped pattern of stars. Now we gain evidence, let us further suppose, that the person is hallucinating with respect to two stars, e.g., when the gaze is fixed on a piece of the sky covered by a dark cloud, he continues to "see" two stars, related to each other in such a way as to make it perfectly intelligible that when two actual stars were so related and placed in appropriate proximity of the S-shaped pattern of stars, one would see an 8-shaped pattern.

It seems highly plausible that the percipient's visual experience of an 8-shaped pattern of stars is a composite visual experience, part veridical (with respect to the S-shaped pattern) and part non-veridical (with respect to the two hallucinatory stars). The seeing of the 8-shaped pattern would seem to consist of two elements, the seeing of the S-shaped pattern and the seeing of the two stars.

According to the epistemic account of perceptual experience, a veridical experience will be the acquiring of true beliefs about the (visual) environment, and non-veridical experience the acquiring of false beliefs. Now in our example, the percipient's total experience, that of seeing an 8-shaped pattern of stars in region r_1 (let us say) will involve the false belief that there is an 8-shaped pattern of stars in that region; the belief is false since there is an S-shaped pattern. Of course the percipient does acquire some true beliefs in this situation, for he also believes that there are *some* stars in this region and there are some there.

However the epistemic account cannot construe the "seeing of the 8-shaped pattern of stars" as part veridical and part non-veridical for that claim would entail two propositions:

- (1) That in seeing the 8-shaped patterns of stars in a certain region, the percipient was acquiring a (true) belief q_1 that there was an S-shaped pattern of stars in that region and a (false) belief q_2 that there are two other stars in close proximity and
- (2) That the belief, p , acquired when the percipient sees an 8-shaped pattern of stars in region r_1 , is a *conjunction* of beliefs q_1 and q_2 (and perhaps some other belief q_3).

The epistemic account implies that these propositions are true, given that the total experience of seeing the 8-shaped pattern is construed as having two *parts*, one veridical and the other non-veridical. If perceptual experience is nothing but the acquiring of beliefs, a part of an experience must be the acquiring of a belief, and to be a part of the whole, the whole must involve a belief that is a conjunction of the belief acquired in the part, together with the beliefs acquired in the other parts.

But each of these propositions (1) and (2) is patently false. It just isn't the case that the percipient would acquire the beliefs q_1 and q_2 , and even if he did, it just isn't the case that belief p is a conjunction of q_1 and q_2 (either of these two alone or of these two with any other belief).

To avoid misunderstanding, let me emphasize that I am not merely claiming that the perceiver doesn't make explicit or conscious judgments, assertions or acts of assenting with respect to the propositions q_1 and q_2 but, more strongly, that he does not acquire these beliefs. This emphatic statement is to allay the claim that one can acquire beliefs without making explicit intellectual or

conscious judgments. And it just isn't true that whenever a person sees in a region of the sky an 8-shaped pattern of stars that person acquires two beliefs, that there is an S-shaped pattern of stars and that there are two other stars, so situated as to turn the S-shaped pattern to an 8-shaped pattern. For one thing it may not be evident to the perceiver whether what turns the S-shape into an 8-shape, is two extra stars or three extra stars. It might just be barely plausible then when a person acquires a belief that there is an 8-shaped pattern of stars in a region of the sky, he also acquires a belief that some of these stars form an S-shaped pattern, but it is certainly not the case that he acquires the belief that *two* extra stars are so placed as to make the 'S' into an '8', nor even the belief that in the region of the sky there is an S-shaped pattern of stars and two other stars. Even if the perceiver does believe that some of the stars fall in an S-shaped pattern, he is unlikely to have any belief about the specific number of the other stars (whether two or three or four).

Moreover the perceiver may see an 8-shaped pattern without believing that a sub-set of the stars form an S-shape for not all tokens of the "8-shape" have a part of the token that is S-shaped. And the perceiver knowing this fact may, even in the situation in which the 8-token has as one of its parts an S-shaped object or set of objects, not acquire the belief that the 8-shaped pattern he is seeing does contain an S-shaped pattern as one of its parts.

However the point could be made even more strongly. For we could choose a different example, one in which the 8-shaped pattern of stars (or lights) is composed of a veridical 2-shaped pattern of stars and several other hallucinatory stars. In this example the 8-shaped pattern is a slightly queer one, usually an 8-shaped pattern does not contain a 2-shaped pattern as one of its parts, but we all know of slightly queer tokens of numerals and alphabetical letters (or if some of us don't, just consult some of the experts designing machines to read). Where it is well known that, in general, an instance of an 8-shaped pattern does not contain an instance of a 2-shaped pattern as one of its parts

(in contrast to the fact that, in general, an instance of an 8-shaped pattern does contain an instance of an S-shaped pattern) then in those rare cases in which the 8-shaped instance does contain a 2-shaped instance as one of its parts, the perceiver is most unlikely to acquire the belief that the 8-shaped instance contains a 2-shaped instance. It is possible of course, but in general most unlikely.

II

To forestall the possible objection that these examples are merely the result of idle fancy, I propose to continue the discussion in terms of an actual psychological phenomenon. In experiments with children who form eidetic images R. N. Haber describes cases³ in which a child is shown a picture of a ship sailing on waves beneath several low-flying birds and clouds. The child forms an eidetic image (a vivid image which the child continues to "see" after the picture is removed) of the picture which is then superimposed on a second picture, one of an apparently unstructured set of lines. As a result the child reports seeing a picture of a bearded man's face. Because the original picture and the apparently unstructured lines are so designed as to form a composite picture of a bearded man's face, it appears to be a plausible explanation of the eidetic child's "seeing the bearded man's face" that he experiences an image that is a composite of two experiences, one veridical and the other non-veridical.⁴

Because in this situation we are considering objects that are images or representations⁵ of other objects, let us take the respective event-types (or object-types) represented by the images to be x , y , and z respectively and let us use the symbol " $r(x)$ " for the description "a representation of event x " or "an x -representation." And let us furthermore distinguish different instances or tokens of the type " x -representation" by the symbols $r_1(x)$, $r_2(x)$, etc.

To explain the child's having the belief (either the false or potential belief⁶) that a representation of an event of type z is in front of him at the time, it seems necessary to suppose that he is both "see-

³ R. N. Haber, "Eidetic Images," *Scientific American*, vol. 220 (1969), pp. 36-44.

⁴ Haber, *ibid.*, p. 44.

⁵ Obviously for experimental purposes the use of pictures and drawings will make for greater ease, convenience and control for the experimenter. But evidence is cited by Haber (p. 43) that the children do form three-dimensional eidetic images so there is reason to believe that one could construct a corresponding composite visual experience with three-dimensional parts.

⁶ It is necessary to speak here of potential belief and not actual belief, for in the case of eidetic images the child in general believes the image to be an image of a previously seen picture, and does not believe it to be a picture.

ing" the eidetic image of a ship sailing, a_1 , which is an $r(x)$, and seeing the unstructured set of lines, b_1 , which is an $r(y)$. We have two representation-instances a_1 and b_1 (an x -representation and a y -representation respectively) which, *when combined in a certain way*, comprises a z -representation c_1 . The two representation-instances a_1 and b_1 are parts of c_1 . By virtue of this fact a_1 besides being an x -representation, is part of a z -representation and more specifically, is an instance of that type of part which complements that type of part of which b_1 is an instance.

Let us symbolize these facts in this way:

- (i) a_1 is an $r(x)$ and is a part of an $r(z)$, i.e., it is a $p[r(z)]$; more specifically a_1 is a certain type of a part of $r(z)$, a p_A type. That is, a_1 is a $p_A[r(z)]$.
- (ii) The other token b_1 is the complementary part of $r(z)$, i.e., b_1 is a $p_A^*[r(z)]$.
- (iii) By adding subscripts to such terms " $r(x)$ " and " $p_A[r(z)]$ " we can form terms to apply to individual tokens of the appropriate types, e.g., $r_1(x)$, $p_A[r(z)]_1$, etc.

Thus we can say of the tokens a_1 , b_1 and c_1 the following:

$$\begin{aligned} a_1 &= r_1(x) = p_A[r(z)]_1 \\ b_1 &= r_1(y) = p_A^*[r(z)]_1 \\ c_1 &= r_1(z) \end{aligned}$$

It is not true that if a_1 is any part of a z -representation and b_1 is any other part of a z -representation, then a_1 and b_1 combined will form a z -representation. The two parts must complement each other. That is, a_1 must be a specific part of a z -representation (p_A) and b_1 must be the complement of that part (p_A^*), i.e., b_1 is a p_A^* . It clearly won't be the case that if a person believes that there are two parts of a z -representation in front of him, he will believe that there is a z -representation in front of him. He would also have to believe that the parts complement each other.

In those cases therefore where one's visual experience consists of two parts one hallucinatory and the other veridical, we will have to describe the experience as the acquiring of a belief that p where the propositional content p can be subdivided into two sub-beliefs, q and r which respectively are the beliefs acquired in having the hallucinatory and veridical sub-parts of the experience. But in order for the epistemic account to work, the sub-beliefs would have to be of a very special sort. In the case where the veridical experience is of some lines

on a paper and the hallucinatory experience an eidetic image of a ship sailing on a moderately cloudy day, then the required belief concerning the lines on the paper would have to be *not* of the form:

"there is a group of lines on paper in front of me,"

but rather:

"there is a group of lines on paper which in conjunction with such-and-such-another group of lines would make a representation of a bearded man's face,"

where it is also believed that the group of lines would in other circumstances constitute a representation of a ship sailing on a cloudy day. But this raises difficulties for the epistemic account, since, as the above discussion was intended to bring out, the propositional content of this belief makes reference to the group of lines contained in the eidetic image. And it would seem that the percipient could only see the first group of lines as part of a bearded-man's-face-representation, if the second group of lines (or the visual-image analogue of a group of lines, i.e., the "material object" of seeing) is presented to the percipient. It is difficult otherwise to make intelligible why the percipient should have *that* belief about the (otherwise) unstructured group of lines in front of him. And even more telling against the epistemic account, we have oversimplified in stating one of the perceiver's sub-beliefs as "there is a group of lines on paper which, in conjunction with such-and-such-another group of lines, would make a representation of a bearded man's face." For this oversimplification suggests that the defender of the epistemic account could make the comeback that the above phenomenon could be rendered intelligible by supposing the hallucinatory part of the experience to be (nothing but) the acquiring of the belief that "there is such-and-such-another group of lines in front of one." But this comeback is not possible for as we have also seen, the perceiver's belief won't in general be about "such-and-such-another" group of lines but instead is likely to be in the form of "*that* group of lines," where the percipient is prepared to explain further what he means in terms of the group of lines being the same group of lines as that constituting the representation-of-the-ship-sailing. The point is that only a specific sort of group of lines (*that* one) will do the appropriate job of complementing the first unstructured group of lines. Alter just slightly some of the curves in the ship-sailing

representation and it is no longer "the sort that will, in conjunction with the first group, make up a bearded-man's-face-representation."

Obviously if we are talking about physical objects, states and events that represent other physical objects, states and events, e.g., TV screens, cinema screens, photographic plates etc., then clearly individual states that separately represent one sort of physical state may be combined to form a composite state that is an image of something quite different from the first two states represented.

In the case of visual experiences, as described in the eidetic image examples, it does appear that one plausible way of describing what occurs is by saying that a person has a complex visual experience representative of some physical state of affairs P_3 , but which is composed of two parts, one of which represents, or is characteristic of, a physical state of affairs of type P_1 , and the other represents, or is characteristic of, a physical state of affairs of type P_2 . One is assuming that the visual experiences have at least two important but very different sorts of features:

- (a) those features internal to the experience themselves whereby, characteristically they represent features of normal "observable" states-of-affairs and
- (b) those features of being, e.g., representative of such and such a physical state of affairs.

According to this view, it is a necessary but not sufficient condition of experiences having features of type (b) (i.e., being representing or "imaging" experiences) that they have features of type (a).

But if, on the other hand, we attempt to analyze visual experience in terms of the acquiring of beliefs (actual or potential), it is difficult to see how we could give a different interpretation to this phenomenon of joining two separate images or states to form a composite image that has a different representing function. The most plausible approach that this sort of account could adopt would be to make use of the distinction between the belief-state considered as a psychological state that the person is in and the propositional content of the belief. Given that a person who has a belief that p is in some psychological state then that state has two sorts of features: (a) internal intrinsic features of the state itself and (b) the feature of having a certain

propositional content, e.g., p . But in order to make good use of this distinction we shall have to suppose that the perceiver is at least capable of being aware (in some sense) of two types of features of his belief-states:

- (a) propositional content
- (b) internal features whereby a particular belief-state has the propositional content specific to it.

But then it is open to the sense-impressionist to claim that these internal features are in the special case of *perceptual* belief the features he ascribes to sense-impressions.

Conclusion

In this paper I have argued that, whatever the merits of the epistemic account in explaining standard hallucinatory and illusory experiences, it cannot account for those perceptual experiences that are composites of veridical and non-veridical elements. For to do so would require the elements to be also the acquirings of perceptual beliefs and situations can be described in which the perceptual beliefs acquired in perceiving the elements do not fit together to make up the set of beliefs (nor an appropriate sub-set of beliefs) acquired when one has the composite perceptual experience.

The two examples used, the figure-8 pattern of stars and the eidetic image case, have been described as containing hallucinatory elements. Since it may be objected that the eidetic-image is illusory rather than hallucinatory, it ought to be pointed out that nothing in my argument depends on the hallucinatory nature of the experience. What is important rather is the non-veridicality of the experiences which, on the epistemic account, is explained in terms of the falseness of the beliefs (actual or potential) acquired. And as far as the feature of non-veridicality is concerned illusory experiences will do as well as hallucinatory experiences. I am inclined to classify eidetic images as hallucinatory because in such cases at the time the child "sees" the image there is nothing there with any of the features discerned in the image. However if by other criteria the image should be classified as illusory rather than as hallucinatory, nothing in my argument will be affected.

VIII. THE INCOMPATIBILITY OF PERCEPTION: A CONTEMPORARY ORTHODOXY

STEWART CANDLISH

A WELL-KNOWN literary critic wrote recently:

One thing that language can do better than any other form of communication is to liberate the individual by setting his imagination to work. If twenty people read the sentence, 'Autumn had come to the woods, and a thin white mist hung between the wet tree-trunks,' twenty separate imaginative pictures will be formed. They will relate to a central core provided by the author, but they will be individual. And so on. In a novel, each reader imagines the physical appearance of the characters, hears the timbre of their voices, pictures to himself the physical setting of the action. In the cinema, by contrast, we all see what the camera puts in front of us. Only one imagination is at work—the director's.¹

Apart from the shortcomings of style so notoriously to be found in the writings of literary critics, a philosopher will see in this passage a gross albeit common error, namely the equation of imagination with the ability to form mental pictures (and their counterparts for other sensory modalities). This error is now so well known that I need only mention it. But another feature of the passage may go unremarked. To see it, let us temporarily suppose with Mr. Wain that exercising the imagination does consist in producing mental pictures. If it is then to be true that when we see a film our imaginations are not at work, seeing something (its being a film is irrelevant) must exclude the formation of mental pictures.

Now in committing himself to this view Mr. Wain would receive not criticism but support from contemporary philosophers, even from those who would reject the mistake just mentioned. Indeed, while some psychologists have thought imaging and perceiving compatible, Anthony Manser writes as though expressing some kind of received philosophical wisdom² in *The Encyclopedia of Philosophy's* article on Images (vol. 4, p. 135):

Mental images have no location and are not related to public visual space; it is useless to ask a subject, as some psychologists have done, to project his mental image onto a screen, for it is impossible to look at the physical world and contemplate an image at the same time.

The baffling implausibility of such a remark in the face of what is known of eidetic imagery indicates that Manser meant only the ordinary rather inadequate visualising which is all that most of us can manage. Yet even with such a restriction, Manser's incompatibility thesis is still unqualified, by no means obviously true, and unargued for in his article.

Manser's views seem to have been influenced by Sartre, and if we turn to *The Psychology of Imagination* (Paris, 1940; English trans. New York, 1948, Methuen paperback edn. 1972) we find that Sartre too holds that imaging and perceiving are incompatible:

But to say "I have an image of Peter" is equivalent to saying not only "I do not see Peter," but also "I see nothing at all". (1, 1, 4: 1972 edn., p. 13)

As long as I am *looking* at this table I cannot form an image of Peter; but if the unreal Peter rises before me all of a sudden, the table which is before my eyes disappears, leaves the scene. (3, 4: *ibid.* p. 138. Cf. also 1, 2, 2: *ibid.* p. 25)

If we are not so pedantic as to take the first remark entirely at face value, Sartre's view is fairly clear, though there is some uneasy wavering between *seeing* and *looking at* to which I shall return. I shall call it the strong thesis. One can read into Sartre's book an attempt to support the thesis by argument when he lists in its early pages the contrasting characteristics of perception and mental imagery. His discussion is marred by some impenetrabilities of style; and by an unhappy tendency to contrast not imaging with perceiving but mental images

¹ John Wain, "The New Puritanism, etc.," *Critical Quarterly*, vol. 14 (1972), pp. 7-18.

² I know of only one modern philosopher who expressly holds a different view on this matter, and that is Brian Smith in *Memory* (London, 1966), p. 154. He acknowledges drawing his opinions on the matter from James Ward's *Psychological Principles* (Cambridge, 1920) and G. F. Stout's *A Manual of Psychology* (London, 1913).

with physical objects—which is a mistake, since the proper objects for comparison are the vertically adjacent members of the two series:

seeing	→ perception (percept)	→ thing seen
imaging	→ mental image	→ thing imaged. ³

This confusion vitiates his attempt to sketch the second of four features of the imaginative consciousness (two are not relevant to the present matter), its quasi-conceptual nature: I do not have to study its objects to determine their properties since nothing is there but what I put there, whereas there is always something fresh to be learned from my perceptions. But even if we eradicate muddle from the exposition of this contrast between imaging and perceiving, at best it will prove only a far weaker thesis than that of their total incompatibility in consciousness: namely, that visualising what one is seeing is pointless since visualising can give one nothing more that is seeably true of an object than one has already in perception, and this we could readily agree to without any special analysis of imagination. Sartre's isolation of a third feature of the imaginative consciousness will also prove only a far weaker thesis than he himself maintains. All consciousness, he says, is intentional, it "posits its object." Perception posits its object as existing and present, whereas imagination either does not posit it as existing or posits it as non-existent, absent or elsewhere. Any incompatibility between perceiving and imagining guaranteed by this claim arises only when the same object is proposed for both forms of consciousness, not between perceiving and imagining in general, and even then the object would have to be known to be the same one. I shall discuss this new, weak thesis later since it is of some interest.

Sartre's contrasts between imaging and perceiving will prove incompatibility only if conjoined with another thesis: that what he calls a consciousness, of one of the three forms he distinguishes (perception, conception, imagination), completely fills the consciousness of any one person at any one time, or at least the consciousness appropriate to a particular sensory mode. It is not entirely clear what this means, but it anyway does not seem to be

true. Why Sartre should think it true, for it does seem to lurk behind what he says, I do not know; I can only suspect the influence of Descartes's opinion that the soul is indivisible. But this is not an essay in the history of ideas.

I know of no one else besides Manser and Sartre who expresses the strong thesis; and we have seen that they do not demonstrate its truth. Are there any reasons for supposing it false? The most obvious reason is simply that it conflicts with experience. Perhaps Sartre and Manser lose their vision completely when visualising, but other people do not. Psychological quirks should not be presented as *a priori* universal truths. Certainly there is, commonly, some difficulty about visualising and perceiving at the same time, but this can very plausibly be attributed to the general problem of attending to two different things at once. Thus some of the plausibility of the strong thesis derives from the fact that in it no clear choice is made between "X is looking at A" and "X is seeing A" when selecting a candidate for incompatibility, and the former particularly does not tell us exactly what the man is doing. Is he concentrating his attention on the sight of A (in which case perhaps visualising as well will be very difficult), is it merely that A is in front of his eyes and he is aware of its presence, or is it that A is in front of his eyes and he is more or less unaware of its presence? But even in the most favourable case for the strong thesis, the temptation to suppose it true arises mainly from the fact that we tend to consider only one kind of example: a man looking at one scene with total concentration and trying to visualise another totally different scene. It is worth noting that one can reasonably be requested, say, to look at a tree and visualise a bird perched in it. The claim that this request is obeyed only by stopping perceiving the tree and visualising it instead is hardly plausible. Secondly, the proponents of the strong thesis have restricted their claims entirely to vision.⁴ If we allow, as we surely must, that there are mental images proper to other sensory modes, we can ask whether any perceiving, no matter in what sensory mode, is incompatible with any imaging, no matter in what mode-correlate. Nothing that Sartre says suggests that he is *entitled*

³ Even if we doubt whether the members of the middle column should be granted the status of *object* (whatever hangs on this), at least the structure of the table is correct.

⁴ Indeed, all of the forms of the incompatibility thesis I know of are framed with respect solely to vision, which is why I have found it most convenient to discuss the general topic by means of visual examples. I am fairly certain that everything I say applies to hearing as well, but have not considered other sensory modes. But I see no reason for supposing that my conclusions cannot be generalised.

to restrict his incompatibility-claim to the one sensory mode; it looks more as if he is committed to supposing that if I hear a dog around the corner then I cannot visualise its being there. This is an unwelcome conclusion, and we may reasonably expect an explanation from the incompatibilists how their thesis is to be restricted in scope. One could, for example, try claiming that necessarily we see things different from those we hear, as the proper objects of sight and sound differ; but such a claim clearly leads us into major controversies. Anyway (just to allay some doubts which may be lingering from my first objection to the strong thesis), if one were to hold *attention* to be the sole source of any incompatibility, as one may be tempted to do, one's account would face precisely this difficulty of explaining the restrictedness of the incompatibility. Further, such an account would deprive the incompatibility of that specialness that its upholders appear to want to claim for it, as it would now seem that the incompatibility is of the same kind as that claimed by "One cannot attend to two parts of a picture at once." Thirdly, while it is true that many people who are, e.g., day-dreaming do not appear aware of what is going on around them, the fact remains that a sudden visual stimulus can generally capture their attention. On the strong thesis this is wholly inexplicable, for presumably the only way we could commence perceiving again would be voluntarily.

In short, the strong incompatibilist thesis is in flagrant conflict with the facts, and must be abandoned. Why have people subscribed, then, to a view so obviously mistaken? Of several reasons, the one of most immediate concern is that the thesis is easily confused with another incompatibilist view by no means obviously mistaken which I shall call the moderate thesis. The moderate thesis is that one cannot visualise what one is now seeing (or looking at). It has been expressed by Ilham Dilman in *Analysis*, vol. 28 (1967-68), p. 96:

... if I do actually see a star I cannot have an image of it. I cannot visualise what I am actually looking at ... my seeing X in my mind's eye, when my eyes are open and I am staring into space, logically presupposes that X is not before my eyes—or if it is, that I do not see it.

and by Roger Squires in *Mind*, vol. 77 (1968), p. 62:

One cannot visualise what one is now seeing because

it is one of the conditions for visualising a scene that one should not be looking at it.⁸

If we ignore the references to looking as well as seeing, as bringing in the already-dismissed factor of attention, we get a very plausible claim which seems, unlike the strong thesis, to fit the facts of experience. Yet it is clear that far more than a mere empirical claim is being made. Both Dilman and Squires maintain that the moderate thesis is *a priori* true. Now it is not an obvious *a priori* truth, so one can expect to be shown reasons for believing it. It might be claimed, e.g., following Ryle, that imaging is a species of pretending and that pretence logically excludes the pretended reality; from which the truth of the moderate thesis would follow. But the Rylean premise of this argument is notoriously doubtful and in any case, Dilman and Squires do not attempt to prove their thesis with such rigour; instead, while they mention pretence, they merely say that imaging is *like* pretending in excluding the imaged or pretended reality, and we are invited somehow to intuit the aptness of the comparison. What are we to say if such intuition fails us? If we are not already prepared to be convinced by the aptness of the comparison then unless it is made out in compelling detail, it will leave us unmoved. We cannot take as demonstrated a claim that X is like Y in respect of P on the basis of knowing only what Y is like. But neither author makes out his case in detail. I am not protesting against the method of comparison in philosophy; on the contrary, I think it useful. But I am protesting against its facile misuse. It is not appropriate to attempted proofs of controversial theses, but only to the elucidation of concepts of which we already have some grasp (for only on such a basis could a comparison be established) but which for some reason are resistant to the formation of informative equivalences.

How, then, can we set about establishing the truth of the moderate thesis? I shall not try to answer this question for I believe the moderate thesis to be, despite its plausibility, mistaken; and I shall instead argue that only a still weaker thesis can be maintained. The claim we are to consider, then, is that seeing X is incompatible with visualising X. One trouble with the claim is this: that I may see X without recognising it as X, and visualise something which I am in fact, though unknowingly, seeing. Suppose, e.g., that you are giving me a

⁸ One can find the moderate thesis in *Zettel*, if one thinks that Wittgenstein intended §§621 and 623 to be taken together and generalised.

present and I am expecting a bottle of champagne. What you hand to me, I take to be a bottle of an inferior wine got up to look like a champagne bottle; disappointed, I visualise your handing to me real champagne. To my shame, I discover that you have after all given me the real thing.

There is an obvious objection to supposing that this is a counter-example to the moderate thesis, namely that the objects of imaging and perceiving are different. What I visualise is your giving me a bottle of champagne, certainly; but what I *see*, despite my error, is your giving me a bottle of an inferior substitute. This objection brings with it too many controversies for us to be dismayed by it, although it could be used to argue that the moderate thesis is defensible if the intentional use of "see" is substituted for the material use. We need not bother, however, to involve ourselves in these controversies about intentionality, for we can admit independently the objection's essential point, which is that it is more plausible to claim some kind of incompatibility between thinking one is seeing *X* and visualising *X*. For ease of exposition I shall discuss only cases where one does see what one thinks one sees; but this is inessential and my points will hold even for cases where one is wrong. I shall term this new incompatibility-claim the weak thesis of the incompatibility between *judgmental seeing*⁶ and visualising.

I speak of judgmental seeing rather than belief or knowledge because I may of course visualise something that I am informed of by some means other than seeing it for myself. There are still further qualifications to be made if the weak thesis is to be defensible, and these are first that a restriction has to be placed upon the objects of visualising and seeing, and secondly that a reference to time has to be included. The objects of visualising must exclude simple physical things and include only situations. Thus I see no reason why I cannot see you, say, walking, and simultaneously visualise you, say, sitting. Similarly, I see no reason why when seeing you performing some action I should not at that time visualise your doing the same thing tomorrow. These restrictions mean that the only viable incompatibility-claim is one concerning visualising and judgmentally seeing the same token-situation. The criteria of identity of different token-situations are not here my direct concern. I

should think any usable ones to be very rough and ready, but some constraints on their formation will become apparent later.

This, then, is the defensible form of the weak thesis. Something like it has, so far as I know, been considered by only one philosopher⁷ (Deitsch, "Visualizing," *Mind*, vol. 81 [1972]) and he discusses only knowingly-seeing and not judgmentally-seeing. Three questions arise. Is the thesis correct? Can we show it to be correct, or incorrect? If it is correct, what is the explanation of the incompatibility? I shall not try to answer the first two questions directly but instead, assuming provisionally the truth of the thesis, shall try to answer the third question and in doing that show indirectly that the first two can receive only heavily qualified affirmative answers.

Let us start by looking at what Deitsch himself has to say on the subject:

... why is it impossible to visualise that which one knows to be the case through sight? Perhaps the only answer possible is that "A is visualising S" entails that A does not know that S is the case through sight ... (pp. 114-115).

Even if we relax *knowing through sight* to *judgmentally seeing* we do not get much help here. Deitsch is claiming that the incompatibility is a logical one; but again it is not obviously so and he does nothing towards demonstrating that it is. So we must manage on our own. I shall provide no neat label for the kind of incompatibility we shall see exhibited, for there is no *neat* incompatibility to be discovered.

To get at the complicated truth of this matter, let us employ a simple distinction between two different relations mental images may have to their objects. The first is a causal relation, analogous to that which a photograph bears to what it is a photograph of. Although the causal relation between image and its object has been disparaged recently (*cf.* Dilman, *op. cit.*), I do not think much reflection is needed to see its importance. One has only, say, to look into a room and then later involuntarily recall its appearance, in order to be aware of the relation's rôle. We can, by virtue of our capacity to acquire and retain knowledge, know the causal origin of an image without inference or the need to trace any intermediate

⁶ This is purely a technical term. Nothing here hangs on whether judgmental seeing is really seeing, or "judgmental seeing" a sense of "seeing."

⁷ To whom I owe the ideas for the two previous examples, and for the qualifications which they illustrate.

causal process; and we may equally be ignorant of the fact that some given image has the direct causal origin that it has. Although this suggests that I am distinguishing some special class of mental images, which I do not think I am, I can find no better way of referring to an image the object of which is determined by this causal relation as a causal-image. The second is an intentional relation, analogous to that which, say, a child's drawing of its mother bears to the woman concerned. The drawing may not have been done from life (hence the causal relation is lacking) and may look nothing like her, but nevertheless she is its object by virtue of the child's intention. This fashionably stressed relation needs no defence at present. With the same proviso as above, I shall call an image the object of which is determined by this intentional relation an intentional-image. I do not pretend that these relations are always exclusive and distinguishable in practice. Presumably I can use causal-images of *X* as intentional-images of *Y* if I know independently that *X* and *Y* are similar, have not seen *Y* but am asked to visualise it. A familiar activity which employs both causal-images and intentional-images is of course visualising.

Now before bringing this tool to the incompatibility question, first notice one immediate corollary of my remark that the only viable incompatibility claim is that between visualising and judgmentally seeing the same token-situation: that there is no *a priori* bar to visualising a token-situation perceptually identical to the one now judgmentally seen. Qualitative identity between token-situations does not of course guarantee numerical identity.

At this point the natures of both the incompatibility and its incompatible terms should be becoming clear. I cannot have a causal-image of what I am now seeing, judgmentally or not, because for the causal condition to be operative on an image the token-situation of which it is a causal-image must have passed. The closest I can get is to have a causal-image of an immediately prior token-situation which is identical to the present one, and use it as an intentional-image of that present one.

But of course as I do not need a causal-image of

X in order to visualise *X*, even where *X* is a token-situation, if there is some bar upon *visualising* and seeing the same token-situation it is not completely constituted by the cause's having to precede its effect. What else might play a rôle? Well, suppose I am looking at an unchanging situation. I am asked to visualise the token-situation I am now seeing. The easiest, if it is available at all, way to visualise something is to employ a causal-image of it, but this way is precluded by the terms of the case: I can utilise only a causal-image of how things were a moment ago. But to do this requires that one neglect what is normally, and in the case of a token-situation almost inevitably, one's best source of information, namely one's present perception. This is difficult to comply with: partly because it is just hard not to concentrate one's attention on the best means of access to what is anyway supposed to be occupying one's attention, and partly because usually one *ought* not to neglect one's best source of information and it is so tantalisingly present. So if I cannot visualise the token-situation I am now seeing, by using an intentional-image of it (which may be a causal-image of an identical but prior token-situation), this is an impossibility contingent upon something like my strength of will.

This account, despite its strangeness, does explain some otherwise puzzling things. It explains why one should be unsure whether or not one can visualise what one is now seeing: for first the answer to this question depends partly on the answer to the question how far can one attend to two things at once and in trying to answer this for oneself one is continually wondering whether one's attention is really split or only constantly switching;³ and secondly the question is easily confused with the one that has an *a priori* negative answer, whether one can have a causal-image of a token-situation one is now seeing. It also explains why the weak thesis is less easily refuted than the moderate thesis for if seeing is restricted to judgmental seeing one knows one has to visualise that of which one seems to be perceptually aware, which poses the psychological difficulties already mentioned; whereas if one is in fact seeing an *X* without knowing this and one is asked to visualise an *X* one is free to utilise one's past perceptions without finding it

³ What is wondered about here is a perfect example of something whose "analysis oscillates between natural science and grammar" (*Philosophical Investigations*, I, §392). For, say, of any two discernibly separate points on a surface, one can be led to concentrate one's visual attention on one to the exclusion of the other, and (quantitative analogies dominating us) it seems that the available amount of attention must be more greatly concentrated when confined to the smaller area. In this way, it eventually becomes a defining criterion of complete attention that it cannot cover two separate points.

hard to neglect one's sight of the *X* under one's nose or to resist the temptation to reach after the never-quite-available best source of information.

The truth of this matter of incompatibility, then, is none of the neat simplicities which philosophers have put forward. We may sum it up as follows:

- (i) No special incompatibility crosses sensory modes. (This rules out finding an explanation solely in terms of attention.)
- (ii) The incompatibility is not between just any imagery and any perceiving within the same sensory mode, but only between those pertaining to one token-situation.
- (iii) The *a priori* incompatibility is between causal-images of *X* and perceptions of *X* and stems from the causal condition's necessity for causal-imagery.
- (iv) Any other difficulty in visualising *X* and perceiving *X* is contingent and stems from difficulties of attention and quasi-moral pressures, and
- (v) the difficulty of (iv) is heightened when only judgmental seeing is considered.

Two things remain. First, does substituting "look at" for "see" make any difference to our conclusions? Although the verb covers a variety of states, the answer is no. For even when it is given its strongest sense of concentrated attention, attention is the only complicating feature it adds, and this is already taken account of in (iv); while if the sense of "look at" is progressively weakened the compatibility approaches more and more the trivial and obvious one of "*A* is visualising *X*" and "*A* has *X* in front of his open eyes."

Secondly, we must account for the confusions

of previous authors. There are several factors. One is that they have failed to distinguish the different relations possible between an image and its object and thus have talked of visualising in terms appropriate only to causal-images; when such an error is allied with the very real difficulty of attending to visualising objects when those things are also under one's nose and the fact that visualising is often accompanied by loss of focus, then a far stronger incompatibility than is really there can begin to seem justified.

Another factor is a tendency to suppose that one's visual field is entirely occupied by perception or by imagery and to forget cases like visualising an (imaginary) bird on a seen bough; when this error is allied with the very pervasive confusion⁹ between mental image and object of imagination it leads to a fallacious account of the supposed incompatibility: for if a mental image is an object of my imagination, is observed, then clearly I cannot be observing both the image and the imaged if each would fill my visual field—one or the other would get in the way, just as would an object and its picture get in each other's way.

And this confusion between mental image and object of imagination itself stems from a very basic philosophical shortcoming—our constantly evidenced inability to account for any fundamental psychological feature except in terms of itself: for external action in terms of inner action (volitions), for perception of objects in terms of perception of inner objects, for visualising objects in terms of inner perception of their images (*cf.* my note on Hannay's *Mental Images in Mind*, 1975). The original confusing picture is still that of the inner observer scanning the inner screen on which there is room for only one picture at a time.¹⁰

University of Western Australia

Received January 10, 1975

⁹ Gilbert Ryle, *The Concept of Mind* (London, 1949), p. 246. Hidé Ishiguro, "Imagination," *Aristotelian Society Supplementary Volume* 41 (1967), p. 51. Ilham Dilman, *op. cit.*, p. 96. Roger Squires, *op. cit.*, p. 62. Brian Smith, *op. cit.*, p. 158.

¹⁰ I am indebted to Kate Candlish for discussion of the subject of this paper.

IX. LISTENER-ORIENTED *VERSUS* SPEAKER-ORIENTED ANALYSIS OF CONVENTIONAL MEANING

EIKE VON SAVIGNY

I. STEPS IN A LISTENER-ORIENTED ANALYSIS¹

IF one conceives of the persons participating in a speech-situation as being divided into speaker, addressee and listener(s) (not addressed) and if one further facilitates this rough simplification by asking which one among these three holds the authority to determine the meaning of an expression, three distinct directions of analysis present themselves: one analysis of the concept of conventional meaning referring to the speaker, another to the addressee and a third to the listener. The first links the meaning of an utterance to the intentions of the speaker, the second to the way the addressee understands it, and the third analysis referring to the listener asks how the utterance ought to be understood. This last analysis must therefore start from the concept of understanding an utterance (-type). The first step makes this clear and is, generally speaking, not especially hazardous.

Step 1: x has for *y* such-and-such a meaning translates into:

y understands *x* in such-and-such a way.

We need not dwell on this step for long, but shall turn instead to the question presenting itself next: What does the second statement signify?

Step 2: y understands *x* in such-and-such a way translates into:

y has, with respect to *x*, that disposition of behavior which makes up the understanding of *x* in such-and-such a way.

This second step is correct only in case the use of the functor in the explanation is justified; if, in other words, there exists for every understanding of a sign one and only one disposition of behavior that makes up that understanding of a sign. The analysis suggested here presupposes such a functional relation. To introduce an example:

Y understands the flashing of the brights of one driver at another driver coming from the opposite direction and waiting to turn left as an invitation to this second driver to make the turn.

If this predicate is validly attributed to a person, the following statements about his behavior ("laws of manifestation" of the disposition) will hold:

(1) If he flashes his brights at a driver in such a situation, he will wait until the driver has completed his turn.

(2) If he flashes his brights at a driver in such a situation from a considerable distance, he will slow down if necessary.

(3) If a driver he has flashed his brights at in such a situation does not complete his turn, he will shake his head, gesticulate or flash his brights again; he will hesitate before driving on.

(4) If in such a situation the brights are flashed at him and the road has only one lane in either direction, he will complete his turn.

(5) If somebody on the second of two opposite lanes flashes his brights at him in such a situation, he will first turn only into this lane.

(6) If somebody flashes his brights at him in such a situation and collides with him in completing his turn, he will get angry.

(7) If the brights are flashed at him in such a situation, he will express thanks with a wave of his hand upon completing his turn.

(8) If he is waiting behind someone not turning left although being invited by such a light signal, he will blow his horn.

(9) If he is a pedestrian about to cross the side road which a driver invited by such a light signal wants to turn into, he will wait.

If these and numerous similar statements about the behavior of a person hold, it will be said that this person understands the flashing of the brights at a driver waiting on the opposite lane to turn left

¹ I think that the following analysis is an adequate interpretation of the ideas of Wittgenstein as sketched in the "Philosophical Investigations". Since my paper is not an attempt at a well-documented interpretation of Wittgenstein, this note should be understood as a warning that I do not think I developed the analysis by myself. In the following it is merely made explicit.

as an invitation to make the turn; and a statement that someone understands this sign in such a way is a claim that these and numerous similar statements hold for him. These reflections win plausibility by imagining characteristic deviations in the behavior of a person; in such a case it usually would be said that he understands the sign differently from normal usage. For example, if there is no reason to assume that a certain person has rather perverse notions of morality, then one will conclude from the fact that he flashes his brights at a driver about to turn left from the opposite lane but accelerates at the same time that he understands this light signal as warning not to turn left. One would also guess that the remainder of the nine statements above do not hold for him, but instead other characteristic statements; this, in effect, would be the content of the proposition that he understands this light signal differently.

Up to now we have only discussed what it means for a certain person to understand a certain sign in a certain way. In the third step, the analysis exchanges the notion of meaning relative to a single person for that of meaning relative to a convention.

Step 3: x , in the convention C , has such-and-such a meaning

translates into:

It is correct, in the convention C , to understand x in such-and-such a way.

Now wouldn't it be rather nearer to intuition to say that x has such-and-such a meaning if all or most people understand x in such-and-such a way? There is of course an element of truth in this; if any utterance (-type) in a language is supposed to have a certain meaning, then all or at least most of the people have to understand it accordingly. This, however, is not enough, for it is not analytic that people understanding an expression differently from the majority understand it "wrong" or "not at all." But this is just what one must be able to say if he wishes to speak of "the" meaning. It is not only the case that an arrow has the meaning of pointing in the direction of its head for most of us, it quite simply has the meaning (in our civilization) of pointing in the direction of the arrow-head. And this means that somebody understanding an arrow as pointing in the opposite direction does not only understand it differently, but also wrongly or not at all. He makes a mistake, and our analysis must allow us to express this. Of course,

beyond this the task remains of clarifying the concept of "correctness" used here.

An informal replacement operation in the third step with the help of the second yields that "It is correct, in the convention C , to understand x in such-and-such a way" translates into "It is correct, in the convention C , to have, with regard to x , that disposition that makes up the understanding of x in such-and-such a way." Applied to our example of the light signal: "It is correct, in Western Germany, to understand the flashing of the brights at a driver waiting on the opposite lane to turn left as an invitation to complete the turn" translates into "It is correct, in Western Germany, to have the disposition characterized by the laws of manifestation (1) to (9)."

But what does it mean to say that it is correct to have a certain disposition? This is an unclear point in the analysis; pre-empting its clarification we will say: "It is correct to have a certain disposition" means: "It is correct, in situations to which the laws of manifestation of the disposition refer, to exhibit the behavior indicated by those laws of manifestation." Beyond certain doubts as to the justification of these operations, they have not added to the content of the first three steps. What is still missing is the clarification of the concept of correctness mentioned above.

If we want to understand correctness in terms of conventions, we obviously cannot take recourse to explicit formulations of such conventions. Not only do most languages lack these, this lack is also a basic feature of natural languages. The way out is thus to shift the focus of our analysis from conventions to the groups in which such conventions hold. For any given convention C , we will call them C -group.

Step 4: It is correct, in the convention C , to exhibit the behavior B in the situation S
translates into:

It is taken for granted to be correct, in the C -group, to exhibit the behavior B in the situation S .

And now for the concept "is taken for granted to be correct."

Step 5: It is taken for granted to be correct, in a group, to exhibit the behavior B in the situation S

translates into

(a) Members of the group seldom openly depart from B in S ;

- (b) if they deviate, they are subject to sanctions on the part of the other members of the group;
- (c) these sanctions are generally accepted.

This explanation for the implicit observance of a rule is generally acceptable and holds not only for rules of language. In our society, for example, the implicit rule holds that we greet acquaintances in the street; this means nothing else than that one usually does this (at least if observed), that one receives minor punishments if one does not (e.g., being gossiped about), and that one accepts these punishments (for example, by apologizing). Watching the newscast on TV on the other hand does not yet seem to have become an implicit rule. Although a lot of people do so, one does not incur sanctions by abstaining; or if one meets disapproval, one does not have to accept it. One could reply, for example, that one's arduous studies do not leave enough time to sit before the "tube"; by this one rejects the sanction.

These three conditions for the observance of a rule are adopted from Hart.³ Condition (a) demands, as he does, that there must be general conformance to a regular pattern of behavior; condition (b) is Hart's condition (1) that deviations must incur criticism; condition (c) is a more general rendering of Hart's condition (2), according to which deviations are accepted as the reason for criticism of the deviation—a specific case of accepting a sanction. Hart's condition (3) treats the "internal aspect" of rules observed in a group. I am unable to see any reason for such an additional condition. Hart's examples are only criticisms of deviations, and therefore do not go beyond his condition (1) (our condition (b)). The "internal aspect" is covered by the other conditions.

Now this explanation has its weak points, because the concepts it operates with can be stretched and are generally somewhat vague. What exactly is "seldom" or "generally"; what are sanctions specifically; under which circumstances must they occur so that someone can be said to be subject to them; how does one accept them? The best approach is to not regard this explanation as a definition of the expression "the rule is observed", but as an enumeration of the circumstances essential to a rule's being observed by a group, even though this enumeration lays claim to completeness.

According to the observations of the author, in

Western Germany the light signal of our example does not only have the mentioned meaning for a great number of drivers here, it simply has this meaning. This implies that the behavior described in the laws of manifestation (1) to (9) and in many similar statements is regarded as correct by the group of drivers and pedestrians in this country in the above sense. For each of these laws of manifestation, therefore, the explanation of the observance of a rule is to be applied. For law (1) this would mean:

- (a) Someone flashing his brights at a driver waiting on the opposite lane to turn left generally waits for him to complete his turn.
- (b) Anyone acting differently in this situation is subject to sanctions (e.g., heated use of the horn and verbal abuse).
- (c) Such a person will generally accept these sanctions (e.g., by waving his hand).

To be sure, the fact should not be overlooked that sanctions can not only be employed against false understanding of the light signal as reflected in behavior, but also, for example, against someone who is suspected of having flashed his brights as an invitation to turn left, but then tricked the complying driver. This ambiguity is reflected in the ambiguity of the sentence: "A driver flashing his brights at a car waiting on the opposite lane to turn left, and slowing down, has to wait for it to make its turn." According to the suggested analysis this is the standard form of a rule of meaning, but of course—and more naturally, too—it could be seen as one of the rules for fair drivers. These two cases may be readily distinguished: The statement expresses a rule of fairness if an argument about it is a moral argument (*sit venia verbo*), and it expresses a rule of language if an argument about it is an argument about the meaning of the light signal. An argument is a moral argument only if the meaning of the light signal is not contended; in other words, only if there is consent on a sufficient number of other rules of the same type (e.g., rules (2) to (9) mentioned above and many similar ones). An argument about the meaning of the light signal arises only if a sufficient number of these rules are in dispute. In this case so many rules are contended that there is no uniform use of language with respect to the sign, and therefore a moral argument cannot be expressed by using this sign.

³ H. L. A. Hart, *The Concept of Law* (Oxford, 1961), pp. 54–56; cf. pp. 87–88, 96, 99–101.

We can now say provisionally: A rule of behavior that relates to an utterance is a rule for the meaning of this utterance only as an element of a large set of mutually compatible rules of behavior also relating to this utterance and together fixing its meaning; a rule deviating from these is a rule of meaning only as an element of another large set of such rules fixing a different meaning of this utterance. A rule is a moral rule only with reference to such a set. Whether such a rule, considered in isolation, is a linguistic or a moral rule is therefore a meaningless question; the reason for its common treatment as a moral rule lies in the presupposition that the meaning of the term used is fixed.

Within the framework of this analysis rules of language are implicitly observed rules, i.e., speech behavior is rule-observing behavior. This does not mean that first there is a rule and secondly a type of behavior complying with it; rather there is only behavior. This behavior has certain characteristic traits which we express by saying that it follows a rule. The rule itself is a subsequently discovered means for describing such behavior. Knowledge of the rule is needed only if one wants to judge whether behavior complies with it; ability to express the rule is no prerequisite for following it. We all follow a colossal number of rules we do not know, among them the rules of our native language. Most of us hardly know a handful of these, and most of the rules themselves are probably not even known to anyone. This does not disturb the native speaker; he follows them just the same, which means nothing but that his speech behavior has the described characteristics.

II. LEWIS' SPEAKER-ORIENTED ANALYSIS

While it is possible to regard the listener-oriented analysis as a combination of Wittgenstein's concept of understanding with Hart's concept of convention, Lewis' basic idea can be regarded as a combination of Grice's concept of meaning with his own concept of convention. This concept radically deviates from that of Hart; it can be approximately characterized in the following way:³

Anyone trying to achieve an objective, or one of several possible objectives (possibly arranged in an order of preference), may have several alternatives of action. These may be more or less adequate for

attaining the desired goal. Their degree of adequacy may depend on the actions of other persons involved in the same situation, striving toward the same end and confronted with the same problem. If two persons *A* and *B* want to meet in New York and for some reason only Times Square and 42nd-and-Broadway can be considered as a meeting place, *A* has the alternative of going to Times Square or to 42nd-and-Broadway. Times Square is the preferable alternative if *B* goes there too, and likewise, 42nd-and-Broadway is preferable if *B* should also turn up there. *B* is confronted with the same but symmetrical problem. Both will achieve their common objective only if from the four possibilities one of the two favorable ones is realized. If *A* can't even guess at *B*'s behavior with any degree of probability and vice versa, both can only trust their luck in choosing one of the two places. Of course, there can be certain circumstances that give *A* grounds to think *B* will go to 42nd-and-Broadway; in such a case he will also go there. Such circumstances could be: *A* has arranged with *B* to meet there ("agreement"); *A* knows that it will be somewhat easier to find *B* at 42nd-and-Broadway than on Times Square ("salience" of the better alternative); or *A* and *B* had already met at 42nd-and-Broadway last time ("precedent"). These circumstances could by themselves make *A* hope that *B* is coming to 42nd-and-Broadway, a hope which motivates *A* to go there.

But a significant additional factor comes into play here. If *A* and *B* both know the other to know the circumstances and if each considers the other to be more or less clever, each will expect the other to expect him to come to 42nd-and-Broadway, so that mutual assumption of ascription of cleverness will provide an additional motivation. (How far this chain continues is unimportant.) *A*'s expectation of *B*'s coming to 42nd-and-Broadway and *B*'s symmetrical expectation—combined with the mutual knowledge of mutual expectation, with the mutual assumption of cleverness, and with the knowledge of this assumption—are sufficient rational ground for *A*'s and *B*'s going to 42nd-and-Broadway in order to meet. It is irrelevant that the expectations were once caused by certain initial circumstances; subsequent meetings on 42nd-and-Broadway are motivated by the complex of expectations. If the meetings turn into a regular pattern of behavior as a func-

³ D. K. Lewis, *Convention* (Cambridge, Mass., 1969). The most important passages on convention in general are on pp. 42, 45, 52-59, 68-70, 84.

tion of the expectations, then this regularity is a convention in Lewis' sense: "For me, a convention is a regularity in behavior produced by a system of expectations." (P. 118.)

If language participants—for any historical reason whatever—mutually expect an expression to be honestly uttered with a certain intention (something to be meant by it), and if the described complex of mutual expectations in this respect actually exists, then it is reasonable for the language participant to fulfill this expectation by using the expression with this honest intention, since in this way he achieves the common objective of all language participants. For Lewis the standard intention is not only that the addressee should (in the case of statements) believe that *p*, but it also involves the speaker's belief that *p*—in other words, that he is honest. If all behave in this way, addressees will usually believe such utterances. The common purpose of all speakers in making such utterances, according to Lewis, is to communicate beliefs. Meaning in a language, then, is what is meant according to a complex of expectations, that is, in accordance with conventions.

Lewis' concept of convention is doubtlessly highly interesting, and what he has made explicit may certainly be called "convention". But the question remains, whether this concept actually functions as it is supposed to for the case of meaning in language, i.e., whether language conventions are Lewis-conventions. I would doubt this. I do not see how one can ask for the measure of knowledge, cleverness and common objectives characteristic of a group following a Lewis-convention among language participants as a prerequisite for stating the existence of a language. The attempt at explication by Lewis has one pervading feature: It is *reasonable* for a group to speak a common language; i.e., to implement a *common purpose* all *act rationally* on the basis of common *knowledge*.

Since the language participants *act* qua language participants, the utterance resulting from an intention is chosen as the basis of the analysis. This is why Lewis comes close to Grice, who provides the means of restricting the analysis to those utterances that can be reported by using the verb "to tell," since both information and orders are carried or accompanied by intentions. This decision is risky to his analysis, because not all utterances and perhaps not even a large part of them are carried or accompanied by intentions having any connection

whatsoever with the meaning of the utterance.

Since rational actions require *purposes*, Lewis searches for and finds them for language. Only his concrete approach in the course of a discussion of signal systems is truly persuasive with respect to this point.⁴ In such systems, signalling actions and the reactions of the receiver are coordinated in such a way that the common purpose of the signaller and signal receiver is achieved. The purpose must be fixed quite well in the concrete situation and the signal system must be adapted in an unmistakable way to this special situation with its special purposes. In such a case it is permissible to talk about purposes that language participants try to achieve together and for the sake of which they coordinate utterances and reactions conventionally; in these cases it is plausible to fix the meaning of the signals by referring to the desired purposes. But if participants in a fully developed language follow certain aims with their utterances, these are normally not purposes for which habits of speaking are stabilized (so that they could play the role in fixing the conventional intention and, therefore, the meaning). The invention of a purpose for which a whole language group uses an entire language is a task of a different order of magnitude; here such formulations as "being able to control one another's beliefs and actions, to some extent, by means of sounds and marks" (p. 181, cf. p. 194)—let alone "communication" (pp. 49, 177, 194)—pale into ciphers. Whatever can be found as a contribution of language to the preservation of homo sapiens is in danger of being subjected to attempts to stylize it into a commonly pursued purpose of the language group.

Since the choice among alternatives based on certain *information* is essential to rational action, the analysis calls for common knowledge of the language participants, and this should be knowledge of the expectations regarding the intention of an utterance and the expectations regarding these expectations. Everyone knows the expected intention of an utterance, everyone expects this knowledge from others, and he expects that they expect the knowledge from himself, and so on. This seems quite a formidable amount of knowledge attributed to the language participants. But it degenerates upon closer scrutiny. In the beginning Lewis speaks explicitly about the beliefs of language participants.⁵ Taken somewhat literally this is an absurdity hardly worth discussing. As he dwells on the

⁴ Ibid., pp. 122-130.

⁵ Ibid., p. 41.

problem, Lewis withdraws from this explicit concept of knowledge in the course of an intelligent and detailed discussion.⁶ In effect, the knowledge of conventions is explained as the competent use of conventions, i.e., the ability to act in a specific case in a way that could be expected if one had the knowledge and used this information correctly. But obviously, the picture would be more pleasing if it were possible to really show the possession of explicit knowledge of the language conventions among language participants and the use of this knowledge, rather than having to refer to the fact that they, in effect, act *as if* they had such knowledge, *but for the fact* that one cannot plausibly assume swift or unconscious steps of reasoning for the average language participant.

In order to make the information effective, all language participants have to show some measure of *cleverness*, e.g., observing Bayes' rule and avoiding errors of multiplication. This cleverness is illustrated by the following passage:

"The Welshmen in parts of Wales use Welsh; each uses Welsh because he expects his neighbours to, and for the sake of communication he wishes to use whatever language his neighbors use." (P. 49.)

Rational thinking of this kind is most conspicuous where steps of reasoning occur; therefore, Lewis'

most detailed explication is kept in this form.⁷ But I concede that this manner of demonstration has didactical reasons; essential to the understanding of Lewis' conception regarding the cleverness of his language participants is the following quotation:

"The intention with which I [use a signal] can be established by examining the practical reasoning that justifies me in [using] it. I need not actually go through that reasoning to have an intention; actions done without deliberation are often done with definite intentions." (P. 155.)

Hence it is no prerequisite that any actual reasoning and calculation have taken place; rather, it is sufficient that we are able to interpret the action as one possibly resulting from rational calculation.

This is the key-word to our question, whether language conventions are Lewis-conventions. Use of language cannot be described in terms of Lewis-conventions; rather these are attempts to reconstruct the use of language by a group in a way that makes it possible to imagine it as consisting of purposeful actions of a number of people working toward a common goal, and whose members possess the necessary knowledge for coordination, and, in general, use this knowledge rationally. And the concessions necessary in this case just show the lack of plausibility of such an interpretation of language.

University of Munich

Received July 18, 1974

⁶ Ibid., pp. 61-68.

⁷ Ibid., pp. 52-57. The expression "actual reasoning" occurs here repeatedly.

X. PSYCHOLOGICAL EGOISM

W. D. GLASGOW

PSYCHOLOGICAL egoism, as traditionally understood, is basically a doctrine of motivation, to the effect that human beings are so constituted that each seeks, and can only seek, his own welfare. There are philosophers who would summarily dismiss this doctrine as unworthy of further study. "It is only *prima facie* a doctrine of motivation," they would say. "Closer analysis reveals that the psychological egoist is trying to elucidate a conceptual truth about what motivates human beings. So that when he asserts that any agent only seeks, and can only seek, his own welfare, the egoist should be interpreted as pointing to the logical truth that any intentional action can always be interpreted in terms of wants." This view is seductive, but not necessarily acceptable to all psychological egoists. The following paper attempts to analyse what those egoists who do not succumb to such arguments are trying to say. It will assume that they wish to say something substantive, in the sense that they believe themselves to be putting forward a strongly based empirical generalization about human motivation. Schlick, for example, who is a psychological hedonist, argues¹ that it is not enough to say that in the majority of cases pleasure is our motivation: we must have a law, that is, a description of relationships that fits all cases. In like manner, it will be assumed in this paper that the psychological egoist is concerned with a law that fits *all* cases.

Because egoism of any variety lends itself easily to misunderstanding, it is important that it should be formulated as carefully as possible. I shall try to do this, making use of the work of Thomas McClintock in this field.² I shall then discuss principle egoism, my revised form of psychological egoism, indicating its deficiencies as a description of human motivation.

1. MCCLINTOCK'S ACCOUNT OF PSYCHOLOGICAL EGOISM

There are, for McClintock, two forms of psycho-

logical egoism, namely motive egoism and principle egoism. He thinks it important to make this distinction because he is interested in the relationship between psychological egoism and normative egoism. Our task, however, is to examine the distinction itself as it is made by him, and to consider to what use it can be put in clarifying psychological egoism.

The two theses of (1) motive egoism, and (2) principle egoism are as follows:

- (1) "Everyone is so constituted by nature that his motive in acting always is to maximize his own interests".
- (2) "Everyone is so constituted by nature that he always acts on the principle 'I ought to act so as to maximize my own interests'."

McClintock believes that the two are compatible with one another, in that principle egoism entails motive egoism, though not *vice versa*. The following quotation³ indicates how the author sees the two types of egoism.

"The first asserts that it is our nature to be moved to act by the belief, thought, or awareness that an action of ours will maximise our own interests. It asserts, in other words, that it is our nature to direct our actions to the realization of one specified goal—maximization of personal interest—because desire to achieve it is what moves us to act; so that we choose, when we do, on a basis of this goal or objective, and are thereby moved to act and do act accordingly. We may say, in sum, that it characterizes us as being, by nature, *practical agents* of an egoistic sort. But the first thesis does not assert that in choosing between alternative actions we are in doubt as to what we ought to do, or are seeking an answer to the question "What *ought* I to do?" It asserts only that we are in doubt as to what *to* do, or are seeking an answer to the question "What *am* I to do?" It is precisely in this respect that the second thesis differs from the first, that the second goes beyond the first. For the second thesis asserts, not merely that we are, by nature, practical agents of an egoistic sort, but that we are, by nature *normative* practical agents of an egoistic

¹ Moritz Schlick, *Problems of Ethics* (New York, 1939), p. 41.

² "The Egoist's Psychological Argument," *American Philosophical Quarterly*, vol. 8 (1971), p. 79.

³ *Op. cit.*, p. 79f.

sort; that is, that it is our nature always to act on the principle "I ought to act so as to maximise my own interests." It asserts, by implication, everything asserted by the first thesis with respect to our motive and corresponding objective in choosing and acting. But in addition to this, it asserts that, in choosing between alternative actions, we are in doubt *fundamentally* as to what we *ought* to do, and only *derivatively* as to what we *are* to do; and consequently that *fundamentally* we decide, or judge, what we *ought* to do, and then *derivatively* choose the corresponding alternative and act accordingly. Therefore, the second thesis entails the first, though not *vice versa*, and is compatible with it."

This distinction which McClintock draws between motive egoism and principle egoism is very puzzling. The motive is the same for both theses, *viz.*, the maximization of personal interest. Both represent men as being practical agents of an egoistic sort, but in principle egoism the agent is represented as something more, namely he is also a normative practical agent. McClintock explains the meaning of this by considering the situation where the agent is confronted by alternative courses of action. In these circumstances, the motive egoist is in doubt only derivatively as to what he is to do, whereas the principle egoist is in doubt fundamentally as to what he *ought* to do. What McClintock does not make clear is the meaning of the distinction between what we might call fundamental doubt and derivative doubt. In what follows I shall attempt to clarify his position. As a result of this I shall modify McClintock's formulations of motive egoism and principle egoism, and conclude that principle egoism is, in relation to human nature, the only possible form of psychological egoism.

2. FUNDAMENTAL DOUBT AND DERIVATIVE DOUBT

One possible interpretation of this distinction might run thus: in motive egoism the egoistic end is never in doubt, and can never be doubted. The motive egoist can experience only what might be called derivative doubt, which presumably is doubt about the best means to achieve the end. This interpretation of motive egoism suggests that by contrast the principle egoist's fundamental doubt *does* have to do with his egoistic end. But what possible doubt could he have about such an end? The doctrine states that he always acts, and indeed can only act on the principle "I ought to act so as to maximize my own interests." Is it possible that the principle egoist could recognize

that something other than his own happiness has intrinsic worth? Is it possible for him to recognize the morality of self-sacrifice, and yet remain a psychological egoist? Many philosophers would doubt this possibility, arguing that if an agent recognized the worth of a course of action, then it must be logically possible, and indeed empirically possible, for him to choose that course. If, however, even one person chose to follow the course of self-sacrifice, then principle egoism could be negated. For it states that everyone *always* acts on the egoistic principle. The oddness of McClintock's description of principle egoism will be discussed later. For the moment, let us conclude that the author, when he talks about the principle egoist's fundamental doubt, is not suggesting that the egoist has any fundamental doubt about his egoism.

Yet the question still remains: what does McClintock mean by fundamental doubt and derivative doubt? A second interpretation of this distinction involves looking more closely at the formulations of motive egoism and principle egoism. First, consider motive egoism:

"Everyone is so constituted by nature that his motive in acting always is to maximize his own interests."

One of the points which McClintock stresses is that the motive egoist is a practical agent but not a normative practical agent. This distinction he does not make clear. What he does suggest is that the practical agent can choose and act from the motive of self-interest, but that he has not the capacity to decide what he *ought* to do. Yet according to the formula of motive egoism, the agent not only chooses a course in accordance with self-interest, he chooses a course which maximizes his self-interest. But surely a person who is conscious of the necessity to maximize his self-interest is conscious of possible alternative actions, and if so he must act, when he does, in terms of a norm. McClintock obviously does not want motive egoism to turn into principle egoism, so it must be concluded that the phrase 'to maximize his own interest' is too strong to be used in the formula for motive egoism. If the concept of maximizing interests is omitted, then motive egoism reads as follows:

"Everyone is so constituted by nature that his motive is always to do what is in his own interests."

But the fact that the phrase "to maximise his own interests" had to be omitted for the sake of con-

sistency, indicates a basic deficiency in motive egoism. For the phrase points to a not uncommon human experience, namely that of trying to do the best thing for oneself in a complex situation. Such an experience involves weighing up the pros and cons of the various alternatives, and of deciding finally in favour of one of them. It is the picture of a rational being making up his mind in accordance with a norm, but it is not, and cannot be, a picture of an agent acting according to motive egoism.

Rather, motive egoism suggests that a man can be aware of his own interests, but cannot feel an obligation to act from them: he does not act on principle. How then does he act? One of the difficulties here is that of language. What words does one use to describe the behaviour of a being who is not, by our standards, fully rational? One could say that he is lazy, or impulsive or careless, to account for the fact that he chooses an alternative that is in his interest, but not necessarily in his best interest (McClintock's derivative choice). Such adjectives could be used, perhaps, but only descriptively, and not evaluatively. It does not seem appropriate to evaluate morally a creature seemingly capable only of concern for the immediate present. But is that not too strong? Motive egoism, it could be argued, does not condemn an agent to his present interests. What, after all, do we mean by "present"? And are there not degrees of rationality? So assuming motive egoism, if an agent could be aware of long-term interests, no doubt he could choose (derivative choice) a long-term alternative in preference to more immediate action. And if he became aware that he had made a bad choice, he could not feel any remorse or sense of guilt, for he is not, according to motive egoism, acting on principle. But perhaps in those circumstances he could feel regret?

In motive egoism, then, the distinction between what the agent does, and what he should do, cannot be recognized. For the agent's motive is not of the form "I want to do what I *think* is in my own interests," nor of the form "I want to do what *seems* to be in my own interests." Both these forms imply that the agent is conscious of a possible gap between what he does and what he should do, and this is not allowed for in the doctrine. The motive rather is of the form "I want to do what is in my own interests." Motive egoism, in fact, does not allow sufficient rationality to the agent for the gap

between "is" and "ought" to be recognized. When, however, the rationality of choice is recognized, the egoist will realize that principle egoism is his only refuge.

3. THE LOGIC OF PRINCIPLE EGOISM

Principle egoism is formulated by McClintock⁴ as follows: "Everyone is so constituted by nature that he always acts on the principle 'I ought to act so as to maximize my own interests'." This formulation is very odd. According to it, the agent *always* acts on the principle "I ought to act so as to maximize my own interests." But surely the nature of practical principles is such that it is empirically possible on occasion to break them. So in principle egoism, the agent may see the principle sometimes as a constraint upon his conduct. He may yield to such constraint, and may subsequently perform actions not in his own best interest. When this happens, it is obvious that the agent is not conforming to principle egoism, at least as described by McClintock. It should therefore be reformulated in the following way:

"Everyone is so constituted by nature that insofar as he acts on any principle, he acts, and can only act, on the principle 'I ought to act so as to maximize my own interests'."

What, then, is the status of this new formulation of principle egoism? It is, in accordance with our assumption at the beginning of the paper, an empirical thesis with the status of a law of nature. Both the motive egoist and the principle egoist are scrutinizing human nature and reaching a conclusion about human motivation. But what the principle egoist is suggesting with his theory is more involved. He seems to be saying in effect "I am not only aware of other human beings and the nature of their motivations, I am also aware of myself as an agent, a being capable of initiating acts in accordance with the egoistic principle, and I am aware that there are other beings like myself in this respect, that they are agents too." In other words, principle egoism involves knowledge of two roles, the role of the agent and the role of the spectator. But it is important to remember that the thesis of principle egoism is psychological in that it begins with the phrase "Everyone is so constituted by nature . . ." Psychological egoism, in this form, sees men as objects to be studied, yet reminds

⁴ *Op. cit.*, p. 79.

us that they are also agents, capable of choosing between alternatives, making decisions, and acting on principle. I shall assume that my reformulation of principle egoism adequately represents psychological egoism.

It is important to note the form of the principle within psychological egoism: "I ought to act so as to maximize my own interests." As a principle this has peculiarities, but it shares at least one characteristic with moral principles, namely that it applies to actions of mine whether past, present or future. According to it, I ought to maximize my own interests not only in the present situation, but on any future occasion in which use of the principle is appropriate. Also there is the implication that I ought to have used the principle on past occasions, whether in fact I did so or not.

There is, however, one characteristic that it does not share with moral principles. Moral principles are universal in that they apply to other people besides myself. If I believe that promises ought to be kept, then I believe that in certain circumstances, it is not only I who ought to fulfil a promise, but *ceteris paribus* you and he and in fact anybody else ought to as well. Some people might argue that in fact it would be better to reformulate psychological egoism in terms of an egoistic universal principle, namely: "Everyone ought to act so as to maximize his own interests." This, I believe, would be contrary to the spirit of psychological egoism, for implicit in such a principle is the recognition of the value of other human beings. The egoistic universal principle states that each should act so as to maximize his own interests. To believe this is to grant at least that there are other autonomous in the sense of having the power to deliberate upon possibilities of actions, to come to decisions and to act upon them. But it is more than this. One who adopts this principle is advocating to his fellows to adopt a particular way to run their lives. But in believing that anyone, say John, ought to act in a certain way, the egoist is committing himself to the position that if John does act in that way, that action would be acceptable to him, even though it is not in his (the egoist's) best interest. Now this poses a problem to any self-respecting egoist, namely how can he support an action that he exhorts others to do that is not in his own best interests. This, however, is not a problem I wish to deal with here. What I wish to suggest is that if the psychological egoist adopts the principle in the form "Everyone ought to . . .," then he commits himself to the position of regarding every

other agent as one whose interests merit equal consideration with his own.

But the psychological egoist is not *necessarily* committed to giving consideration to others, except insofar as such consideration is in his own interest. Indeed to argue that he is committed seems contrary to his principal concern—that his only duty is to further his own interests. If this is right then it seems that in psychological egoism we must substitute for the egoistic universal principle, the principle: "I ought to act to maximize my own interests."

So the egoistic singular principle becomes as important for the egoist as moral principles for the non-egoist. But it is the one and only possible dominant principle, according to the egoist, and this characteristic makes its logic differ in an important way from the logic of moral principles. To bring out this difference, let us suppose that an intelligent agent, with knowledge of the relevant facts, regularly breaks a moral principle which he professes to hold. At first we (other moral agents) may call him weak-willed, but as his transgressions continue, we may come to suspect that he does not really hold the principle, that in fact he has given it up. When this happens, we would still think of him as a moral agent, but with a different set of principles, and indeed it is possible that even if he does break a moral principle on occasion, he may still claim to hold it, provided he condemns his own actions, and has feeling of guilt. On the other hand, the egoist, observing other agents regularly acting in a way contrary to the egoistic principle, would never come to the conclusion that they have given up the egoistic principle, for that is impossible. Human nature being what it is, people can act only in accordance with the principle. The egoist must therefore explain such apparently non-egoistic behaviour in other ways, *e.g.*, in terms of the agent's ignorance of the circumstances, or in terms of his lack of intelligence, impetuosity or some similar explanation. The professed egoist may also, like the moral agent, condemn his own actions when he breaks his principle, and he may also experience guilt feelings. Yet even if the guilt feelings disappear in time, and even if he ceases to condemn his transgressions, the egoistic principle would still be the principle to which he is committed. In fact, he can never seriously challenge or criticize his principle, in the way that non-egoists believe that moral principles can be criticized. For if the egoist were to criticize it, he could do so only in the light of a non-egoistic fundamental

principle, which by implication he would hold. This, however, it is logically impossible for the psychological egoist to do. He might maintain in fact that even the concept of such a non-egoistic principle is at the best empty and void.

Such is the logic of principle egoism. No doubt to the non-egoist it is a crazy logic. He will be appalled by the egoist's disbelief in the possibility of generous or benevolent acts, as he will be angered by the dogmatic insistence on there being possible only one fundamental principle. His differences with the egoist are not resolvable by discussion or debate, for they spring from differences in basic standpoints. A passage from Iris Murdoch hints at what the egoist's vision of the world is likely to be. Hume, she says,⁵ taught us "to see the world in

terms of contingently conjoined simples, to see it as a totality of ultimate simple facts which have no necessary connection with each other. Insofar as we imagine the world does contain necessities, and that real connections exist between these simple elements, this is merely the result of habit and custom, which are themselves the work of nature. It is only in reflective moments that we can see the ultimately disjointed character of the world. It is habit which gives to us, according to Hume, both our objective material world, and our moral world." So the reflective egoist, perceiving the ultimately disjointed character of the world, perceives also man's essentially nonsocial character and as a result formulates psychological egoism.

University of Sheffield

Received November 7, 1974

⁵ In "Metaphysics and Ethics," *The Nature of Metaphysics* (London, 1957), edited by D. F. Pears, p. 113.

XI. PART-OF-THE-MEANING-OF-A-WORD

JAMES F. HARRIS, JR.

PERHAPS it is only fools who rush into a territory so well trodden by the angels; however, I have come to believe that the arguments of some of the anti-intensionalists¹ have ignored some of the important subtleties and multifarious features of language of which speech-act analysis has informed us. Also, the suggestion that pursuing questions about meanings (concepts, ideas) of words is always a waste of time philosophically has the effect of directing us away from a course of philosophical inquiry which might prove very fruitful. I am particularly concerned that questions bearing upon the analytic-synthetic distinction become suspect, e.g., "Is the meaning of 'unmarried' part of the meaning of 'bachelor'?" and "Is 'All bachelors are unmarried' analytic or synthetic?" Questions concerning parts of meanings are very important for the intuitively appealing claims that meanings of some terms are sometimes "contained in" or "part of" meanings of other terms, e.g., "'Unmarried' is *part of* the meaning of 'bachelor'."

I wish to examine what sense can be made for talking about parts of meanings after we have recast what is called "the traditional problem of meaning" within a speech-act framework. The question ought not to be "Do meanings of words have parts?" where we raise the question about *no particular* word at all, just "any word," and thereby commit what Austin calls the fallacy of being about "Nothing-in-particular."² Neither do I wish to reify meanings, *a fortiori* parts of meanings. Indeed, I am intent upon arguing that talk of parts of meanings can be sensible without meanings or their parts *existing* ontologically.

To begin with, the use of the spatial metaphor "part of" unfortunately shifts a question which is really a semantical one into an ontological consideration. And it is quite proper, I think, to claim that it is nonsense to think of meanings as having

parts the ways in which automobiles have parts. Hopefully, the suggestion offered here will skirt the ontological question altogether. Indeed, Austin himself has provided us with the theory which allows us to talk about parts of meanings without having to deal with it on the ontological level. If we can answer "What-is-the-meaning-of (the word) 'x'?" without committing ourselves to some entity, the-meaning-of-'x', then we can do the same for questions about parts of meanings.

The use of interrogatives suggests a way of clarifying the relationship which sometimes exists between the meanings of two different particular words. I will not offer here a general definition of "analytic" or "synthetic" or "synonymy," nor do I think, that such a general definition is necessary to comment meaningfully upon the problem at hand. Whatever we have to say then about the relationships of meanings will always depend upon the particulars.

Consider the assertive uses of the following statements:

- (100) Socrates is a professor.
- (101) Socrates is a male.
- (102) Socrates is a male professor.³

Now it seems obvious that part of what we would want to say about these statements is that (102) logically implies (not simply "gives it to be understood") both (100) and (101). Also, (100) does not imply either (101) or (102), and (101) implies neither (100) nor (102). We can say all of these things without involving ourselves in any ontological dispute about Socrates, professors, males, or meanings since we simply are making claims about the relationships between statements.

If implication is problematic for some Quinean reasons of circularity, the problem can be approached differently by the use of interrogatives. Suppose that you are asked the following questions:

¹ See W. V. O. Quine, "Two Dogmas of Empiricism," *The Philosophical Review*, vol. 60 (1951), pp. 20-43; Nelson Goodman, "On Likeness of Meaning," *Analysis*, vol. 10, (1949), pp. 1-7; and J. L. Austin, "The Meaning of a Word," *Philosophical Papers* (Second Edition), ed. by J. O. Urmson and G. J. Warnock (Oxford, 1970).

² Austin, *ibid.*, p. 58.

³ "Male professor" is a compound predicate, I realize; and however, the use of compounds illustrates the relationship of meanings well, and some compounds must be regarded as completely new substantives.

(103) Is Socrates a professor?

(104) Is Socrates a male?

(105) Is Socrates a male professor?

Now admittedly one would have to know something about the world to be able to give the correct answers to these questions, but we are not concerned here with the correct answers but the relations which exist, if any, among the possible answers.

Suppose that (103) is answered affirmatively (whether correctly or incorrectly). If the answer to (103) is "Yes," does that help answer either (104) or (105)? For (104), answering (103) affirmatively does not help. For (105), the proper response seems to be "partially" or "somewhat." If (105) is understood as asking a conjunctive question

(105A) Is Socrates a male?

and

(105B) Is Socrates a professor?

then both (105A) and (105B) separately must be answered affirmatively in order for (105) to be answered affirmatively. So, if (103) (which is the same question as (105B) is answered affirmatively, we have half of the answer to (105).

Suppose that the answer to (103) is negative (again, whether correctly or incorrectly). If the answer to (103) is "No," does that help answer either (104) or (105)? Again, answering (103) either affirmatively or negatively does not help to answer (104). But if (103) is answered negatively, then (barring equivocation, reference shift, etc.) the answer to (105) is obviously negative also.

Also, if (105) is answered affirmatively first, then the answers to both (103) and (104) are already provided for us.

What all of this suggests is that answers to questions about meanings are the sorts of things which can have parts. Let us see how far this approach will take us forward on elucidation of the relationship between meanings of particular words.

First, we must ask questions about meanings of particular words rather than about Socrates. Consider, then

(106) What-is-the-meaning-of "professor"?

(107) What-is-the-meaning-of "male"?

(108) What-is-the-meaning-of "male professor"?

An answer to (108) (whatever it is) provides us with answers to both (106) and (107). And answers to (106) and (107) separately provide us with partial answers to (108). Whereas an answer to (106) does not help in answering (107) nor does an

answer to (107) aid in answering (106). This obviously suggests some relationship between the meanings of "professor" and "male professor" which does not exist between the meanings of "professor" and "male."

In speech-act analysis, "the-meaning-of 'x'" has properly been dropped in favor of "what-we-mean-by 'x'"; so why not replace "part-of-the-meaning-of 'x'" by "part-of-what-we-mean-by 'x'?" The suggestion I am offering here is that if one does what is required to answer (108), then he does not have to do it again or do something else different or additional in order to do what is required to answer (106) and (107).

Initially, it may seem that it would be possible to syntactically generalize what we have discovered so far to apply to any compound. We cannot. Although (or *because*) we answer

(109) Is Socrates a fake professor?

affirmatively, we cannot answer

(110) Is Socrates a professor?

affirmatively also. There appears to be no syntactical way of generalizing the relationship between (105) and (104) and (105) and (103) and also account for the difference between (105) and (109). There are numerous other adjectives which, when used in compounds, would create the same difficulty as does "fake professor," e.g., "simulated," "phony," "fictitious," "imitation," "feigned"—to name a few. A general syntactical formula, e.g.,

(111) The meaning of 'x' is part of the meaning of the compound 'xy' then will not work where *x* is any noun and *y* is any adjective.

To try to accommodate the difference between (105) and (109), I want to suggest that uses of modifying adjectives (by position) can normally be divided into the following categories.

Adjectives			
Revisors		Qualifiers	
Visitors	Neutralizers	Predicators	Emphasizers
fake	possible	male	genuine
imitation	alleged	black	authentic
counterfeit	putative	red	real
phony	conceivable	smart	true
bogus	imaginable	large	
simulated	potential	small, etc.	
feigned	rumored		
imaginary	supposed		
forged	proposed		
fictitious			
mythical			
fraudulent			
sham			

Revisors have the effect of revising the original noun. Vitiators either void, hollow, vitiate, or completely negate the original noun. A counterfeit ten dollar bill, for example is not a ten dollar bill at all, and neither is a phony doctor a kind of doctor. A feigned headache is not a headache, and a simulated fight is no fight. Neutralizers have the effect of "leaving the question open" about the relationship of meaning between the original noun and its occurrence in the compound. An alleged murderer, for example, may or may not be a murderer. And a rumored love affair may or may not be a love affair. Predicators simply ascribe some predicate to the original noun resulting in a greater degree of specificity. "Male professor" is more specific than "professor," and male professors is a sub-class of professors. Emphasizers, though qualifiers, do not ascribe some new predicate to the original noun, and their use does not result in a greater degree of specificity. Compounds using emphasizees simply repeat the original noun more emphatically or with more emphasis. A genuine ten dollar bill is still just a ten dollar bill, and all ten dollar bills are genuine ten-dollar bills.

The significance of the distinction between revisors and qualifiers lies in the fact that while (111) will not work when y is a revisor, it will when y is a qualifier. The only additional stipulation which must be made for (111) to work is that x be used in a purely referential position (whether it actually refers or not). In

(112) What-is-the-meaning-of "Sunday"?

the formula stated in (111) will not work because neither "sun" nor "day" occurs in a purely referential position in "Sunday."

The next step is to propose a way of deciding whether an adjective is a revisor or qualifier. Although there have been major objections to certain features of Moore's use of the open question test, I suggest that such a test can here be used to identify revisors and qualifiers. The test is being used here just as a test; it is not being offered as a definition. Where x is any individual, N is any noun (or substantive), and A is any adjective, then if

(113) x is a AN

is taken as true and it is still an open question (i.e., it is not trivial or silly) to ask

(114) Is x a N ?

then we will say that the adjective being substituted for A fails the open question test and is a revisor.

However, if (114) is a closed (silly, trivial) question, then the adjective being substituted for A passes the open question test and is a qualifier. In other words, if by answering (113), one also answers (114) the adjective in question is a qualifier.

The purpose of introducing this distinction can now be made clear. Compounds formed by the use of a noun and a qualifying adjective have meanings which we commonly say are the sort of things which can have parts; whereas compounds formed by the use of a noun and revising adjective we do not commonly talk about as having meanings which are the sorts of things which have parts. Looking back at the relationship between (105) and (103) discussed earlier, we can now say that that relationship is explained by the fact that "male" is a qualifier since it satisfied the open question test. However, since "fake" fails the open question test and is a revisor, the relationship between (109) and (110) is different.

It is important to note that one particular substitution of an adjective may pass the open question test and another substitution of the same adjective may fail. Thus, "male" in "male professor" is a qualifier but "male" in "male aunt" is a revisor. Combinations of revisors cause difficulties which must be handled *ad hoc*.

Since no qualification of the original substantive takes place when the adjective is a revisor, it seems that we would do better to regard the compound formed by using a revisor as a totally new substantive with its own meaning and extension. At least, we should recognize the way in which the meanings of compounds using qualifiers are derivative and parasitic upon the original substantives—a way in which the meanings of compounds formed by using revisors are not derivative or parasitic.

The shift made earlier from asking questions like (103)–(105) to questions like (106)–(108) may suggest that by recasting the questions as is done in (106)–(108), I have changed what really is a problem of reference into a problem of meaning. That is, (103)–(105) suggest that the problem is one of whether to ascribe certain predicates to a particular individual and has nothing to do with the meanings of those predicates themselves. But consider the problem raised by (109) and (110). Since there is a *logical* difficulty involved in answering both questions affirmatively, then there must be more to the problem than simply whether a particular individual just happens to be included in the extension of two different expressions. To assert both

(115) Socrates is a fake professor

and

(116) Socrates is a professor

while not contradictory, still creates a logical difficulty similar to asserting 'The cat is on the mat, and I do not believe it.' The problem cannot be handled simply as a problem of reference but must be regarded as a problem of meaning as well.

Comparisons with other "What is . . ." questions illustrate how the problems at hand transcend simple ontological considerations. Austin thinks that square roots of numbers are unlike meanings of words.

I can ask not only the question, 'What is the square root of 4?' of 8, and so on, but also 'What is the square root of a number?': which is either nonsense or equivalent to 'What is the "square root" of a number?' I then give a definition of the 'square root' of a number, such that, for any given number x , 'the square root of x ' is a definite description of another number y . This differs from our case in that 'the meaning of p ' is not a definite description of any entity.⁴

It seems strange indeed to suggest that answering such a question about the square root of a given number commits a person to the existence of square roots as metaphysical entities. What is called a definite description of a number is simply a description of a definite, specifiable operation for deter-

mining an answer to the question "What is the square root of x ?"⁵ Consider "what is the time of the concert?" To answer this question we need make no commitments at all concerning the existence of time. There is no procedure to determine the meaning of no word at all, but neither is there a procedure for determining the square root of no number at all just as there is no procedure for determining the time of no event at all. However, I fail to see what this has to do with separate and entirely different questions concerning the ontological statuses of time, square roots and meanings as entities.

Finally, we are in a position to see how the above points bear on the analytic-synthetic distinction. If I have succeeded in demonstrating that in one sense it *sometimes* makes sense to talk about meanings having parts, then it will *sometimes* make sense to talk of sentences being analytic. We seem justified in claiming *at least* as much as the following: Where x is any individual, N is any noun (or substantive), and A is any adjective then any statement of the form

(115) If x is an AN , then x is a N

is analytic where A is a qualifier (determined by the open question test). Whether such considerations can be applied to other kinds of statements traditionally regarded as analytic is a question still under consideration.

College of William and Mary

Received October 23, 1974

⁴ *Op. cit.*, p. 60.

⁵ And ' x ' functions here to designate some specific unknown number.

XII. PUNISHMENT WITHOUT OFFENCE

T. S. CHAMPLIN

No matter what is done to an innocent man, you cannot punish him, however much you may make him suffer. This view has been expressed by philosophers as dissimilar as Kierkegaard¹ and Hobbes.² Bradley³ opined that such a belief in the necessary connection of punishment and guilt was typical of "the man of uncultivated morals" and subscribed to it himself with obvious relish. Amongst contemporary philosophers⁴ the view seems to have become actually something of a commonplace and their position is often called logical retributivism.⁵

I

An examination of the authors quoted above (see footnotes 1-4 for references to specific passages) will, I believe, reveal that in each case the same argument is being invoked. It is the aim of this first section to expose a fallacy in this popular argument. What this argument is can be gauged from the following extract from Benn and Peters.⁶

We said at the beginning of this chapter that 'punishment' implied, in its primary sense, not the inflicting of *any* sort of suffering, but inflicting suffering under certain specified conditions, one of which was that it must be for a breach of a rule. Now if we insist on this criterion for the use of the word, 'punishment of the innocent' becomes a logical impossibility. For it follows from the definition of 'punishment' that suffering inflicted on the innocent cannot be '*punishment*'. It is not a question of what is morally justified,

but of what is logically possible. (In a footnote they add) An analogous relation between 'guilt' and 'pardon' accounts for the oddness of 'granting a free pardon' to a convicted man later found to be innocent.

The argument then is that all punishment is $\exists$ OR something⁷ (here, for a breach of a rule); and that where that "something" is lacking (here, no breach of a rule as in the case of an innocent man), there can be no punishment. This argument gets what plausibility it has as a result of failing to distinguish between various ways of specifying that for which a person is being punished. Contrast:

- (i) Punishing a boy for the noise he made in class.
- (ii) Punishing a boy for a noise that was made in class.
- (iii) Punishing a boy for making a noise in class.

Doing either (i) or (ii) entails doing (iii); but doing (iii) does not entail doing either (i) or (ii). Even when no noise at all has been made, it is still possible for some luckless boy to be punished for making a noise in class. The conclusion "There can be no punishment" in the above argument does not follow from "There can be no punishment of any actual breach of a rule" (which would follow from the premisses); and furthermore it is false because there can be punishment *for* a breach of a rule, even when there can be no punishment *of* an actual breach because no rule has been broken.

¹ Søren Kierkegaard, *Purity of Heart* (New York, 1956), pp. 97-98.

² Thomas Hobbes, *Leviathan* (Oxford, 1960), p. 207.

³ F. H. Bradley, *Ethical Studies* (2nd edn., Oxford 1927), Essay I, pp. 26-27.

⁴ E.g., Anthony Quinton, "On Punishment," *Analysis* (1954), repr. in Peter Laslett (ed.), *Philosophy, Politics and Society* (Oxford, 1956), p. 86; C. W. K. Mundle, "Punishment and Desert," *Philosophical Quarterly*, vol. 4 (1954), p. 220; S. I. Benn and R. S. Peters, *Social Principles and the Democratic State* (London, 1959), p. 182; K. G. Armstrong, "The Retributivist Hits Back," *Mind*, vol. 52 (1961), pp. 478-481; J. D. Mabbott, *An Introduction to Ethics* (Hutchinson Univ. Library 1966), p. 136; Peter Winch, "Can a Good Man be Harmed?", *Proceedings of the Aristotelian Society*, vol. 66, 1965-6, repr. in Peter Winch, *Ethics and Action* (Boston, 1972), pp. 197-200; H. F. Pitkin, *Wittgenstein and Justice* (Berkeley, 1972), p. 188.

⁵ After, I think, Anthony Quinton (q.v. footnote 4 *supra*) who spoke of retributivism as being not a moral but a logical doctrine.

⁶ Q.v. footnote 4.

⁷ It is, of course, not true that all punishment is $\exists$ OR something. Some cars are built to take a lot of punishment. Athletes take part in punishing races. Boxers punish each other. It need not make sense to ask "What are they being punished for?" However, I shall ignore this complexity and consider only cases where the punishment is $\exists$ OR something.

II

If the official argument in favour of the view that the innocent cannot be punished is invalid and its conclusion is false, why has the view appealed to so many able philosophers? I believe that the reason is that it has been confused with the subtly different view that it is impossible to punish innocence (i.e., the innocent can be punished, but not for being innocent, not for their innocence). It is the aim of this second section to provide an argument in support of this second claim and at the same time to explain why attempts at stating the relationship between punishment and innocence are likely to give rise to paradox. My argument is of a quite general sort and is not restricted to punishment. In fact, it can be applied to any verb permitting the three grammatical constructions "to ϕ someone for something," "to ϕ someone" and "to ϕ something."

Kierkegaard was, I believe, gesturing at this second claim in the following famous passage⁸ from *Purity of Heart* which has been analysed elsewhere by Peter Winch.⁹

Now and then someone speaks of 'suffering punishment, when one does the Good'. How is that possible? From whom shall that punishment come? Certainly not from God! Is it, then, from the world—so that when in its wisdom the world is mistaken, it rewards the bad and punishes the Good? And yet no, it is not as the word 'world' implies. The word does not mean what it says. It is improperly expressed. For the word 'world' sounds great and terrifying, and yet it must obey the same law as the most insignificant and miserable man. But even if the world gathered all its strength, there is one thing it is not able to do, it can no more punish an innocent one than it can put a dead person to death.

It is possible to take another verb, e.g., "to correct" as in "to correct someone for a spelling mistake," which like "to punish" passes my grammatical test, and to construct a parody of Kierkegaard's reasoning. This will allow me to argue in favour of the paradoxical view (a) that one cannot correct a person for spelling a word correctly (i.e., the correct spelling of a word cannot be corrected), and (b) that nevertheless a person who has just spelled a word correctly can be corrected (i.e., the correct spelling of a word can be corrected).

⁸ Q.v. footnote 1.

⁹ Q.v. footnote 4.

¹⁰ Except in "inverted values" cases, e.g., where a gangleader punishes any member for not doing something nasty each week. Similar "inversion" in spelling examples arises where there are two standards of correctness, e.g., American versus English spelling of "colour." But then these would not be counter-examples.

Kierkegaard, on my view, like many other philosophers, might have seen the first part of this, but would have been led, not unnaturally—but nevertheless mistakenly—to deny the second part.

Imagine that Joe takes spelling lessons with two teachers, Miss Perfect and Mr Power. Miss Perfect is kind and understanding and never wrong or careless; Mr Power is bad-tempered, often careless and sometimes even wrong. Joe tells his mother about some of the things that have taken place in spelling class. She reasons as follows:

Now and then Joe speaks of 'being corrected, when he has spelled a word correctly.' How is that possible? From whom could such correction come? Certainly not from Miss Perfect! Could it be, then, from Mr Power—so that when in his wisdom Mr Power is mistaken, he praises an incorrect speller and corrects a correct one? And yet no, it is not as the name 'Power' implies. The name does not mean what it says. It is not a proper name for him. 'Mr Power' sounds great and terrifying, and yet he must obey the same law as the most insignificant and miserable man. But even if Mr Power gathered all his strength, there is one thing he does not have the power to do, he can no more correct a boy for spelling a word correctly than he can put a dead person to death.

In order to understand what Joe's mother is saying, we need to bear in mind the important distinction between correcting someone for his mis-spelling of a word, and correcting someone for misspelling a word. The former entails that he did misspell the word, the latter does not. Joe's mother is only thinking in terms of the second type of construction. She has noticed the absurdity of Mr Power's explaining to an onlooker that he was correcting Joe for spelling a word correctly in answer to the question "Why are you correcting Joe?" Similarly, we could not understand the referee who explained after the football match that he penalised the centreforward for taking the corner-kick legally; or the father who tried to justify his action by saying that he punished his son for not having done anything wrong;¹⁰ or the king who punished his subjects for being innocent of plots on his life, or who rewarded them for cowardice in battle with medals inscribed with the words "For Cowardice." What makes all of these things unintelligible emerges in objections to them like "But a correctly spelled word isn't a spelling mistake . . . A legally taken

corner-kick isn't a foul . . . Innocence isn't a crime . . . Cowardice isn't a merit in a man."

It is vital, however, to notice that Joe's mother's claim is quite compatible with Joe's having been corrected for his spelling of the word "jeopardy," which was in fact perfectly correct, by Mr Power who did not pay attention to what Joe actually said and who thought, mistakenly, that he spelled it "jea. . . ." So it is possible for Joe's correct spelling of a word to be corrected; and yet his mother was right to insist that you cannot be corrected for spelling a word correctly. The real paradox of spelling-correction, then, is that one cannot be corrected for spelling a word correctly (only for misspelling it) and yet one can be corrected for spelling a word correctly! This astonishing paradox arises from the fact that what appear to be contradictions are not contradictory really. Consider the sentence

- (1) A fine was imposed on John *for illegal parking* for having left his car in the High Street.

It was Bill's tough luck to be fined *for dangerous parking* for the same offence. Bill both was, and wasn't, fined for the same offence—which illustrates how what follows the "for" can be taken in two entirely different ways, because identical words are used. In (1) we have a grammatical construction of the form "to be $\emptyset$ ed for X for Y ." This double "for" construction can be reproduced for many verbs¹¹ taking "for," including "punish for." The double "for" construction can be split up into separate sentences. One can be $\emptyset$ ed for X ,

- (2) A fine was imposed on John *for illegal parking*.

and one can be $\emptyset$ ed for Y ,

- (3) A fine was imposed on John for having left his car in the High Street.

If we concentrate on Y , i.e., on what follows the second "for," we can intelligibly ask whether it was legal or not, even if we accept that (1) as a whole is true. A man can, after all, be fined *for illegal parking* for leaving his car in a perfectly legal spot—perhaps no one realised, or believed, that it was legal at the time of the trial. There may be many a hard luck story of this sort to be told. But if we

turn to X , i.e., to what follows the first "for," there is no comparable question that can be asked about the legality or otherwise of what John has been given a fine for. Indeed, if you try to replace "illegal" by "legal" in (1), you fail to make sense in the same way that fining people for keeping the law lacks sense, except in the topsy-turvy situation in which a new law outlaws the old. One can be $\emptyset$ ed for X , and X can be $\emptyset$ ed,

- (4) *Illegal parking* is to have a fine imposed on it. Not just a polite warning, as in the good old days. One can be $\emptyset$ ed for Y , and Y can be $\emptyset$ ed.

- (5) John's leaving his car on the High Street had a fine imposed on it.

Yet John's parking spot may have been quite legal.

If your thoughts are solely directed towards examples like (4), you'll see the point of saying that the idea of imposing fines on *legal* parking is unintelligible. This is why Kierkegaard and Joe's mother speak as they do about punishing and correcting. If your thoughts are solely directed to examples like (5), you'll be anxious to remind us that magistrates are fallible, that we don't always get justice in this world, etc. The fact is that both parties are right; it is only a grammatical illusion that they are contradicting each other. Each side errs when it supposes that its own position excludes the other's.

In conclusion I shall make the same point specifically in terms of punishment. Consider

- (6) John was punished for selfishness for taking two pieces of cake.

Let us suppose that John did take two pieces, but he meant to save one for a sick friend. It is Kierkegaard's point that "selfishness" could not be intelligibly replaced by "unselfishness" in (6). But the fact that unselfishness cannot be punished leads him to deny that John was punished in this example. A hard-headed philosopher can see that John has been punished for his unselfish act in this example and this leads him to deny the point Kierkegaard is making. Each is right in what he asserts and wrong in what he denies. One and the same man *can* speak with the voice of "eternity" and the voice of the "world."

¹¹ E.g., "He was *treated* for cancer for his throat condition," "He was *rewarded* for gallantry for rescuing a comrade," "He was *reported* for professional misconduct for the way he handled his patients," "He was *penalised* for bargaining for allowing his shoulder to come in contact with the goalkeeper."

The Philosophical Quarterly

CONTENTS OF VOLUME 26

No. 102

JANUARY 1976

HUME BICENTENARY NUMBER

THREE QUESTIONS ABOUT JUSTICE
IN THE *TREATISE*

Antony Flew

HUME ON MORALITY AND THE
EMOTIONS

Stewart Sutherland

SIMPLICITY, RESEMBLANCE AND
CONTRARIETY IN HUME

R. J. Hawkins

HUME'S ONTOLOGICAL COMMITMENTS

Wade L. Robison

HUME'S AESTHETICS REASSESSED *Peter Jones*

PHILO'S CONFESSION

W. A. Parent

MIRACLES AND COMMON UNDERSTANDING

P. S. Wadia

HAS THE AUTHORSHIP OF THE *ABSTRACT*
REALLY BEEN DECIDED?

John O. Nelson

MORE HUME BIBLIOGRAPHY

Roland Hall

Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant, Anthony Ellis, and Stephen Read

Published for the Scots Philosophical Club by the University of St Andrews

Annual Subscriptions: Individuals £3.00 (\$8.50) Institutions £5.00 (\$14.00)

Single Issues £1.00 or \$3.00 (Institutions £1.50 or \$4.00)

for orders placed directly with the Journals Manager, Scottish Academic Press
Ltd., 25 Perth St., Edinburgh EH3 5DW, Scotland

MIND — JANUARY 1976

Edited by
D. W. HAMLYN

CENTENARY NUMBER

Articles:

Editorial: A Hundred Years of *Mind*:
THE EDITOR

George Croom Robertson: Editor
1876-1891: A. QUINTON

G. F. Stout's Editorship of *Mind*
(1892-1920): J. A. PASSMORE

Mind under G. E. Moore (1921-1947):
C. LEWY

Gilbert Ryle's Editorship:
G. J. WARNOCK

Mind under the Editorship of Hamlyn:
J. N. FINDLAY

Improvisation: G. RYLE

Book Reviews

Discussions:

Intention and Foresight in Law:
R. LYONS

Morality and Egoism: C. H. WHITELEY

Zombies, Schizophrenics, and Purely
Physical Objects: D. LOCKE

Double Deception: T. S. CHAMPLIN

A Rebuttal of Pollock's 'Refutation' of
Non-cognitivism: F. E. McDERMOTT

On the Recurrence of Things Past:
D. W. THEOBALD

The Justification of Deduction: S. HAACK

Books Received

Published for the MIND ASSOCIATION by BASIL BLACKWELL

Single Issues: £1.25

Annual Subscription: £4.00

ANNOUNCING

The Charles S. Peirce Foundation, Inc.

The Charles S. Peirce Foundation has been established to support research on the work of Charles S. Peirce, educational programs concerning his life and thought, and original research building upon and continuing or correcting his work. The Foundation will assist the National Park Service and the Delaware Water Gap National Recreational Area in providing for the renovation and use of Arisbe, the homestead of Charles S. and Juliette Peirce near Milford, Pennsylvania, for the above purposes and as a national memorial to Charles S. Peirce as scientist, mathematician, logician, and philosopher.

The Foundation has been organized as a non-profit corporation under the statutes of the Commonwealth of Pennsylvania. Application for tax-deductible status is currently pending before the Internal Revenue Service.

The officers and Board of Directors of the Foundation are: Edward C. Moore, President; Carolyn Eisele, Secretary-Treasurer; Max H. Fisch, Gresham Riley, and a representative to be designated from the Department of Philosophy of Harvard University.

Two classes of membership in the Foundation are available, for which the fees are twenty-five dollars for annual membership and five hundred dollars for life membership. Life memberships may also be obtained in five payments of one hundred dollars each year.

Persons interested in becoming members of the Foundation or in contributing to its work should address the Secretary-Treasurer, Professor Carolyn Eisele, 215 East 68th Street, New York, New York 10021. Checks may be made to the Charles S. Peirce Foundation.

Journal for the Theory of Social Behaviour

VOLUME 5 NO. 2 OCTOBER 1975

Edited by P. F. SECORD (Department of Urban Studies, Queens College, New York) and
T. MISCHEL (Department of Philosophy, SUNY at Binghamton, New York) assisted by
R. HARRÉ (Oxford University)

Contents

THOMAS M. OLSHEWSKY
Dispositions and Reductionism in Psychology

BERENT ENÇ
On the Theory of Action

D. C. DENNETT
Why the Law of Effect will not Go Away

KENT BACH
Analytical Social Philosophy—Basic Concepts

LOUIS A. MOFFETT
Art Objects as People—A New Paradigm for the
Psychology of Art

EVERT VAN DER VLIERT
Gross, Mason and McEachern have not Really
Verified their Theory of Role Conflict Resolution

J. RICHARD EISER
Attitudes and the Use of Evaluative Language:
A Two Way Process

Annual Subscription (*two issues*) £5.50 (\$16.50) for institutions and £3.50 (\$10.50) for individuals
Single issues £2.75 (\$8.25)

Published by **BASIL BLACKWELL**

Orders to 108 Cowley Road, Oxford OX4 1JF, England

R. G. COLLINGWOOD MEMORIAL FUND

Robin George Collingwood (1889–1943) was a Fellow and Tutor in Philosophy at Pembroke College, Oxford, from 1912–1935. During this period he published or prepared his most important philosophical and historical works.

The College has decided to establish an annual essay prize named after him and is appealing for contributions from those who would like to express their appreciation of his work. Cheques, made payable to Pembroke College and crossed R. G. Collingwood Fund, should be sent to the Vicegerent, Pembroke College, Oxford, England, from whom further information may also be obtained.

DIALOGUE

Canadian Philosophical Review—Revue canadienne de philosophie

Vol. XIV—No. 3 Septembre—1975—September

Editors—Redaction

François Duchesneau (Université d'Ottawa)

John Woods (University of Victoria)

ARTICLES

On Being Present to the Mind: A Sketch for the History of an Idea

Hume's Theory of the Self Revisited

Queries on Truth—Conditions

On The Third Attempted Definition of Knowledge, *Theaetetus* 201c—210b

Enseignement et endoctrinement

Deux traditions épistémologiques (dialogue)

L'éternel retour chez Nietzsche: une faillite philosophique

Philosophie et violence chez Eric Weil

JOHN W. YOLTON

TERENCE PENELHUM

HUGHES LEBLANC

MAY YOH

OLIVIER REBOUL

JOHN KING-FARLOW & PRUDENCE JONES

DANIEL LETOCHA

JEAN ROY

NOTE—DISCUSSION

Propositions and Adverbial Modifiers

WAYNE A. LENHARDT

COMPTES RENDUS—BOOK REVIEWS

Membership—Cotisation annuelle: \$16.00; Students—Etudiants: \$5.00.

Subscriptions—Abonnements: \$16.00; Single Copies—Le numéro: \$5.00

Association canadienne de philosophie—Canadian Philosophical Association

Suite 46, 1390 ouest rue Sherbrooke (Sherbrooke St. West), Montréal, Québec, H3G 1K2

The British Journal for the Philosophy of Science

Contents, Volume 26, Number 3, September 1975

Articles

- | | |
|--------------|---|
| COLIN HOWSON | — The Rule of Succession, Inductive Logic and Probability Logic |
| PAUL HORWICH | — Grünbaum on the Metric of Space and Time |
| YORICK WILKS | — Putnam and Clarke and Mind and Body |

Discussions

- | | |
|------------------|---|
| ARTHUR THOMAS | — Puccetti on Machine Pattern Recognition |
| J. W. ROXBEE COX | — Mackie on Dispositional Properties |
| JAMES CARGILE | — Newcomb's Paradox |

Review Articles

- | | |
|--------------------|--|
| L. PEARCE WILLIAMS | — Should Philosophers be Allowed to Write History? (Review of J. Agassi: <i>Faraday as Natural Philosopher</i> and W. Berkson: <i>Forces and Field</i>) |
|--------------------|--|

MAURICE FINOCCHIARO—Philosophising about Galileo (Review of D. Shapere: *Galileo: A Philosophical Study*)

Reviews

RUPERT HALL on N. Bukharin (ed.): *Science at the Cross Roads*; CHARLES SCHMITT on Galileo's: *Two New Sciences*, translated by Stillman Drake; MARGARET BODEN on B. Meltzer and D. Michie (eds.): *Machine Intelligence 7* and R. C. Schank and K. M. Colby (eds.): *Computer Models of Thought and Language*

Announcement

Published for the British Society for the Philosophy of Science by
Aberdeen University Press, Farmers Hall, Aberdeen AB9 2XT, Scotland

Editors: John Watkins and John Worrall

Price £2.50 (\$8.00)

Subscription Price £7.00 (\$22.00)

GRAZER PHILOSOPHISCHE STUDIEN

INTERNATIONALE ZEITSCHRIFT FOR ANALYTISCHE PHILOSOPHIE

gemeinsam mit

K. Acham, R. M. Chisholm, R. Freundlich, E. Topitsch und O. Weinberger

herausgegeben von Rudolf HALLER

Board of Consulting Editors: H. ALBERT (Mannheim, BRD)—D. FØLLESDAL (Oslo, Norway)—P. GOCHET (Liège, Belgium)—J. HINTIKKA (Helsinki, Finland)—J. KIM (Ann Arbor, Mass., USA)—F. v. KUTSCHERA (Regensburg, BRD)—K. LAMBERT (Irvine, Cal., USA)—K. LEHRER (Tucson, Arizona, USA)—J. K. MACKIE (Oxford, England)—G. PATZIG (Göttingen, BRD)—M. PRZELECKI (Warsaw, Poland)—W. STEGMÜLLER (Munich, BRD)—P. WEINGARTNER (Salzburg, Austria).

Review Editor & Editorial Assistant: Wolfgang L. GOMBOCZ, Reinhard FABIAN

A major objective of "GRAZER PHILOSOPHISCHE STUDIEN" is to stimulate and to strengthen communication between American and European analytical philosophy. "GRAZER PHILOSOPHISCHE STUDIEN" are published by RODOPI of Amsterdam. One or two volumes (from time to time on special topics) of at least 200 pages will appear yearly. Correspondence concerning editorial matters, manuscripts for publication (in duplicate) and books submitted for review should be sent to Prof. Dr. Rudolf HALLER, Philosophisches Institut, A-8010 Universität Graz, Austria. Business correspondence and subscriptions please address to EDITIONS RODOPI N.V., Keizersgracht 302-304, Amsterdam, Holland.

Out of the contents of volume 1 (1975)

BRENTANO, Franz von: Was an Reid zu loben. Über die Philosophie von Thomas Reid (Aus dem Nachlass)
CHISHOLM, Frederick M.: Individuation. Some Thomistic Questions and Answers.
LEHRER, Keith: Remembering without Knowing.
WEINBERGER, Ota: Wissensaussage und die Unmöglichkeit ihrer Objektivierung.

Further articles by Wilhelm K. ESSLER, Ronald E. BEANBLOSSOM, Terence PARSONS, Ronald SCALES, Michael HOOKER, David W. SMITH, Ulrich MAJER and Rainer STUHLMANN-LAEISZ as well as review articles by Thomas PAXSON, Henri SARLET, and Kurt WEINKE.

the philosophical forum A QUARTERLY

Volume V, No. 3
Marx W. Wartofsky, Editor

- Merrilee Salmon** Representation and Intention in Art
Kai Neilsen Rationality, Intelligibility, and Alasdair MacIntyre's Talk of God
Maurice A. Finocchiaro The Concept of *Ad Hominem* Argument
in Galileo and Locke
Clifford E. Williams Now, Extensional Interchangeability
and the Passage of Time
Paul Fitzgerald Critical Study of *The Language of Time*, by Richard M. Gale.
Serge J. Morin Maine de Biran: A New Dualism
Alphonso F. Lingus The Perception of Others
F. F. Centore Atomism and Plato's *Theaetetus*
David E. Starr The Sixth Sophist
Linda Nicholson Feminism: Reform or Revolution? A Reply

\$10/yr. Individuals/\$20/yr. Institutions
Mail to: Department of Philosophy, Boston University,
232 Bay State Road, Boston, Mass. 02215

Journal of Philosophical Logic

Editor-in-Chief: Bas C. van Fraassen, University of Toronto, Ont., Canada.

Editorial Board: L. Aqvist, N. D. Belnap, Jr., M. J. Cresswell, J. E. Fenstad, F. Fitch, D. Føllesdal, P. T. Geach, C. Hamblin, H. G. Herzberger, J. Hintikka, W. Hodges, S. Kanger, D. Kaplan, S. Kripke, K. Lambert, R. B. Marcus, G. J. Massey, D. Prawitz, N. Rescher, A. Robinson, W. C. Salmon, D. Scott, K. Segerberg, P. F. Strawson, P. Suppes, R. H. Thomason, G. H. von Wright.

Executive Committee: N. D. Belnap, Jr., B. C. van Fraassen, G. J. Massey, N. Rescher.

The journal is concerned with philosophical studies utilizing formal methods or dealing with topics in logical theory. Specifically included are: contributions to branches of logical theory directly related to philosophical concerns; contributions to philosophical discussions that utilize the machinery of formal logic; discussions of philosophical issues relating to logic and the logical structure of language; and philosophical work relating to special sciences (for example linguistics, history of logic, or physics), with emphasis on foundational problems.

Subscription price per volume of four issues Dfl. 100,- / US \$ 40.00, plus Dfl. 15,- / US \$ 6.00 for postage and handling. One volume will be published yearly. Personal subscription price on request. Reduced price for members of the Association for Symbolic Logic Dfl. 40,- / US \$ 16.00 per volume, including postage and handling.



D. Reidel Publishing Company

Dordrecht - Holland / Boston - U.S.A.

Synthese

An International Journal for Epistemology and Philosophy of Science

Editor-in-Chief: Jaakko Hintikka, Academy of Finland and Stanford University.

Associate Editor: Ilkka Niiniluoto, Institute of Philosophy, Helsinki.

Review Editor: Dagfinn Føllesdal, Institute of Philosophy, University of Oslo.

Editorial Assistant: Juha Manninen, Institute of Philosophy, Helsinki.

Board of Consulting Editors: David M. Armstrong, Yehoshua Bar-Hillel, Arthur W. Burks, Robert S. Cohen, Donald Davidson, Hans Freudenthal, Ian Hacking, Jerrold J. Katz, Imre Lakatos, G. Nuchelmans, Wesley C. Salmon, Wolfgang Stegmüller, Erik Stenius, Patrick Suppes, Klemens Szaniawski, Ladislav Tondl, H. Törnebohm, Marx W. Wartofsky.

Synthese publishes articles in all the fields covered by the subtitle. These include the theory of knowledge; the general methodological problems of science; the methodological and foundational problems of the different departmental sciences; those aspects of symbolic logic and of the foundations of mathematics which are relevant to the philosophy and methodology of science; and those facets of the history and sociology of science which are important for contemporary topical pursuits.

Subscription price per volume of four issues Dfl. 100,- / US \$40.00, plus Dfl. 12.50 / US \$5.00 for postage and handling. Two volumes will be published yearly. Personal subscription price on request.



D. Reidel Publishing Company

Dordrecht - Holland / Boston - U.S.A.

377

ETHICS

Since 1890 an International forum for ethical theory, jurisprudence, and social philosophy, **Ethics** discusses the fundamental structures and values of civilized society. The journal is committed to no particular policy, social viewpoint, philosophic style or doctrine, but rather synthesizes the work of many disciplines in order to develop and clarify the bases for personal and social choices. Writers contribute major essays and reviews on such subjects as the nature of freedom, morals and morality, legal justice, the organization of society, and social issues such as violence, equality, and abortion.

Articles in 1975

Holmes Rolston III, Is There an Ecological Ethic?

William Leon McBride, The Concept of Justice in Marx, Engels, and Others

William T. Blackstone, Freedom and Women

Marcia Aufhauser, Guilt and Guilt Feeling: Power and the Limits of Power

Norman C. Gillespie, A Valid Deduction of the Generalization Argument

Published quarterly by The University of Chicago Press

Warner Wick, Editor

Please enter my one year subscription: ☐ Institutions: \$14.00

☐ Individuals: \$10.50 ☐ Students (with signature of professor): \$8.40

Countries other than USA add \$1.00 to cover postage.

Please mail your check or purchase order to **Ethics**, The University of Chicago Press, 11030 Langley Avenue, Chicago, Illinois 60628

AMERICAN PHILOSOPHICAL QUARTERLY MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

In March 1976 we shall be publishing **NEGATION AND NON-BEING** by Richard Gale, which is the tenth in this impressive Monograph series. The price is \$8.00 (£4.00) to institutional subscribers and to non-subscribers; individual subscribers, however, will continue to receive the Monograph as part of their subscription. In past years we were able to supply Monographs to all subscribers to the Quarterly, but this has for the time being become impossible. If your library is a subscriber to the Quarterly, please check whether it is now subscribing also to the Monograph series. It is only in the Monographs that the larger themes of modern philosophy can receive extended treatment, and those who subscribe only to the Quarterly are missing an essential part of the coverage of philosophy which is provided by the journal as a whole.

To enable you to judge for yourself the importance of this series, a complete list of the Monographs is given overleaf. The last issue supplied to institutions on subscription was No. 7 (1973). If you find that Nos. 8 (1974) and 9 (1975) are still missing from your files, we shall be glad to supply them at the quoted prices, and to enter your standing order for all future issues of the series.

PRINTED IN ENGLAND

by William Clowes & Sons, Limited, London, Beccles and Colchester

AMERICAN PHILOSOPHICAL QUARTERLY MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

- No. 1. **STUDIES IN MORAL PHILOSOPHY.** *Contents:* Kai Nielson, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Hāworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$8.00 (£4.00).
- No. 2. **STUDIES IN LOGICAL THEORY.** *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$8.00 (£4.00).
- No. 3. **STUDIES IN THE PHILOSOPHY OF SCIENCE.** *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$8.00 (£4.00).
- No. 4. **STUDIES IN THE THEORY OF KNOWLEDGE.** *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, \$8.00 (£4.00).
- No. 5. **DESCARTES' PHILOSOPHY OF NATURE** by James Collins. 1971, \$8.00 (£4.00).
- No. 6. **STUDIES IN THE PHILOSOPHY OF MIND.** *Contents:* Harold Brown, "Perception and Meaning"; Haskell Fain and A. P. Griffiths, "On Falsely Believing That One Doesn't Know"; O. H. Green, "Emotions and Belief"; Donald F. Henze, "Descartes on Other Minds"; Moreland Perkins, "Meaning and Feeling"; and Alan R. White, "What We Believe." 1972, \$8.00 (£4.00).
- No. 7. **STUDIES IN ETHICS.** *Contents:* Charles V. Blatz, "Mitigating and Meliorating Defenses"; Claudia Card, "Retributive Penal Liability"; Norman O. Dahl, "Is Mill's Hedonism Inconsistent?"; R. L. Franklin, "Recent Work on Ethical Naturalism"; James D. Wallace, "Cowardice and Courage"; A. D. Woozley, "Injustice." 1973, \$8.00 (£4.00).
- No. 8. **STUDIES IN MODALITY** by Nicholas Rescher, with the collaboration of Ruth Manor, Arnold vander Nat, and Zane Parks. 1974, \$8.00 (£4.00).
- No. 9. **STUDIES IN EPISTEMOLOGY.** *Contents:* Dennis M. Ahern, "Hume on the Evidential Impossibility of Miracles"; Robert Almeder, "Fallibilism and the Ultimate Irreversible Opinion"; Jonathan Bennet, "Stimulus, Response, Meaning"; Peter A. French, "'Seeing' and 'Seeing that', 'Observing' and 'Observing that'"; Margaret Gilbert and Fred R. Berger, "On an Argument for the Impossibility of Prediction in the Social Sciences"; David K. Glidden, "Protagorean Relativism and Cyrenaics"; Mark Pastin, "Modest Foundationalism and Self-Warrant"; E. E. Sleinis, "Normal Conditions and Perceivers." 1975, \$8.00 (£4.00).

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Myles Brand
Robert Butts
Roderick M. Chisholm
L. Jonathan Cohen

James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Carl G. Hempel
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Richard Rorty
Jerome Schneewind
George A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 13/NUMBER 2

APRIL, 1976

CONTENTS

- | | | | |
|--|-----|---|-----|
| I. GERALD J. MASSEY: <i>Tom, Dick, and Harry, and All the King's Men</i> | 89 | IV. DORIS OLIN: <i>Newcomb's Problem: Further Investigations</i> | 129 |
| II. JOHN LESLIE: <i>The Value of Time</i> | 109 | V. SIDNEY AXINN and DAVID AXINN: <i>Notes on the Logic of the Ignorance Relations</i> | 135 |
| III. MARILYN FRYE: <i>On Saying</i> | 123 | | |

PAPERS PRESENTED AT THE 50TH MEETING OF THE PACIFIC DIVISION OF THE AMERICAN PHILOSOPHICAL ASSOCIATION (BERKELEY, MARCH, 1976)

- | | | | |
|--|-----|--|-----|
| VI. DAVID LEWIS: <i>The Paradoxes of Time Travel</i> | 145 | VIII. PETER HEATH: <i>The Incredulous Hume</i> | 159 |
| VII. J. O. URMSON: <i>Fiction</i> | 153 | IX. RICHARD WASSERSTROM: <i>The University and the Case for Preferential Treatment</i> | 165 |

© AMERICAN PHILOSOPHICAL QUARTERLY, 1976

PUBLISHED IN OXFORD BY BASIL BLACKWELL

ISSN 0003-0481

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The price *per annum* is twelve dollars for individual subscribers and twenty-eight dollars for institutions. Checks and money orders should be made payable to the publishers, Basil Blackwell & Mott Ltd. Back issues are sold at the rate of three dollars to individuals, six dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, U.K.).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The back cover of the journal may be consulted for details.



I. TOM, DICK, AND HARRY, AND ALL THE KING'S MEN

GERALD J. MASSEY

I. THE POVERTY OF PREDICATE LOGIC

AUGUSTUS DeMorgan ridiculed traditional (neo-Aristotelian) logic for its inability to justify the inference from "All horses are animals" to "All heads of horses are heads of animals," a relational inference easily legitimated by standard predicate logic. Perhaps because of its success within mathematics—all inferences made in classical mathematics receive formalizations that render them valid—predicate logic has been subjected to no scorn like DeMorgan's; at worst its critics have decried its insensitivity to the variety of uses of "if... then" in everyday parlance, a peccadillo compared to the alleged shortcomings of traditional logic. But in fact there are inferences altogether as elementary as DeMorgan's that elude predicate logic, inferences like (A1)–(A3):

- (A1) Tom and Dick carried the piano upstairs.
Dick and Tom carried the piano upstairs.
- (A2) Tom, Dick, and Harry are shipmates.
Dick, Harry, and Tom are shipmates.
- (A3) Tom, Dick and Harry are shipmates.
Tom and Harry are shipmates.
- (A1') Tom and Dick love Mary.
Dick and Tom love Mary.
- (A2') Tom, Dick, and Harry are linguists.
Dick, Harry, and Tom are linguists.
- (A3') Tom, Dick, and Harry are linguists.
Tom and Harry are linguists.

The inability of predicate logic to validate (A1)–(A3) has been masked by its manifest ability to justify the superficially similar arguments (A1')–(A3'). An apologist of predicate logic might claim that (A1)–(A3) can be handled within predicate

logic as follows. As to (A2) and (A3), one first introduces dyadic and triadic *shipmate* predicates governed by meaning postulates (1) and (2).

- (1) $(\forall x)(\forall y)(\forall z)(S^3xyz \supset S^2xzy \ \& \ S^2yxz)$
- (2) $(\forall x)(\forall y)(\forall z)(S^3xyz \supset S^2xy)$

Then one treats (A2) and (A3) as enthymemes, having (1) and (1) & (2) respectively as suppressed premisses.

Unquestionably, this *enthymematic ploy* brings (A2) and (A3) within the explanatory scope of predicate logic.¹ So powerful a strategem is the enthymematic ploy that it can bring *any* intuitively valid argument within the explanatory scope of *any* logic that incorporates *modus ponens*; as the "suppressed" premiss, simply take the associated conditional of the given argument. (This suppressed premiss will qualify as a meaning postulate when the original argument is intuitively valid.) Traditional logic incorporated *modus ponens* in the guise of the so-called hypothetical syllogism, so it could use the enthymematic ploy to validate DeMorgan's argument, just as predicate logic uses it to account for the validity of (A2) and (A3). Contemporary students of logic sometimes express amazement at the obstinacy of traditional logicians who did not immediately renounce their impoverished logic when the incomparably richer logic of Frege and Whitehead-Russell appeared. There is little to marvel at; their resort to the enthymematic ploy made them oblivious to the limitations of their logic, just as it makes modern logicians oblivious to the shortcomings of theirs. The enthymematic ploy is a panacea, not for remedying a logic's deficiencies, but for rendering them invisible.

Even if our tentative defense of predicate logic had not proved to be a recipe for bringing any intuitively valid argument within the explanatory scope of any logic, it would have been unsatisfactory. *Enthymeme* and *suppressed premiss* are

¹ The enthymematic ploy is treated at length in [18].

psychological notions, not logical ones. Despite some murkiness, we know how to use these notions fairly well. In particular, we would not ordinarily saddle with suppressed premisses someone who advanced (A2) or (A3) as arguments, for these arguments stand on their own logical feet. What we should demand of our logic is an account of their validity that is grounded in their *internal* (deep) structures, an account that does not gratuitously invoke "suppressed" premisses where we detect none. Furthermore, we should demand *systematic motivation* for whatever deep structures our logic assigns to these arguments. For example, even though (A2) and (A3) would thereby acquire internal structures that rendered them valid, we should balk at accepting (3) and (4).

- (3) $S^2th \ \& \ (\forall x)(\forall y)(S^2xy \supset S^2yx)$
 (4) $S^3tdh \ \& \ (\forall x)(\forall y)(\forall z)(S^3xyz \supset S^3xzy \ \& \ S^3yxz \ \& \ S^3xy)$

as the respective deep structures of (5) and (6),

- (5) Tom and Harry are shipmates.
 (6) Tom, Dick, and Harry are shipmates.

unless (i) tetradic and higher degree *shipmate* predicates are handled analogously, (ii) "shipmates" is shown to belong to a natural class of predicates all of which are treated alike, and (iii) intuitively valid arguments relevantly similar to (A2) and (A3) come out valid by virtue of their deep structures. The scheme implicit in (3) and (4) for dealing with *shipmate* predicates founders on the reef of (iii). A logic that accounts for the validity of (A2) and (A3) ought to be able to legitimate (A7).

- (A7) All the men there are are shipmates.
 Tom and Dick are distinct men.

 Tom and Dick are shipmates.

The formalization scheme implicit in (3) and (4), however, does not permit uniform treatment of sentences like (7),

- (7) All the men that there are are shipmates.

since that scheme calls for *shipmate* predicates of fixed degree, dyadic, triadic, and so on. But no *shipmate* predicates of fixed degree will allow an adequate translation of (7). For any choice of n

as the degree of the *shipmate* predicate of highest degree in a symbolic rendering of (7), there will be arguments altogether like (A7) whose internal structures will fail to validate them, e.g., (A8):

- (A8) All the men that there are are shipmates.
 $a_1, \dots, a_n$, and a_{n+1} are distinct men.

 $a_1, \dots, a_n$, and a_{n+1} are shipmates.

The cause of this failing is obvious. When formalized as indicated, the first premiss of (A8) says of any n men that they are shipmates; it does not mean that all the men there are, whatever be their number, are shipmates.

Let us call predicate logic, when supplemented by Leonard & Goodman's logic of individuals, *mereological predicate logic*.² Most logicians and philosophers seem to regard mereological predicate logic as a theory that furnishes solutions to certain technical problems in constructionalist metaphysics or as a theory congenial to nominalists and others with an aversion to sets. Lately, however, it has been used by Parsons [21] and by Burge [2] to give a truth theory of English sentences containing mass terms. I will show that mereological predicate logic is more central to our everyday conceptual scheme than common opinion would have it or even than Parsons' and Burge's application would suggest. To do this, I will show that mereological predicate logic furnishes deep structures that resolve problems about phrasal conjunction and group quantification that have perplexed contemporary linguists. *Inter alia*, arguments (A1)–(A3) and (A7) can all be proved valid in mereological predicate logic. In addition, the theory of phrasal conjunction and group quantification presented herein has potentially far-reaching consequences for philosophical analysis.

II. THE ONTOLOGY OF INDIVIDUALS

Leonard & Goodman developed their *logic* or *calculus of individuals* as an adjunct to set theory so that constructionalist philosophers could deal with predicates like "shipmates" with a clear philosophical conscience. The fundamental distinction drawn by this calculus is that between *classes* (or *sets*) and *individuals* (or *wholes*):

² *Mereology*, the logic of individuals, was first developed by Leśniewski whose principal papers appeared in Polish. It was formalized and made available to English-speaking logicians by Leonard & Goodman in their classic paper [17]. Goodman's book [9] contains an informative presentation of the logic of individuals or, as it is usually called, the *calculus of individuals*. A concise, axiomatic development of the calculus of individuals can be found in my book [19].

The concept of an individual and that of a class may be regarded as different devices for distinguishing one segment of the total universe from all that remains. In both cases, the differentiated segment is potentially divisible, and may even be physically discontinuous. The difference in the concepts lies in this: that to conceive a segment as a whole or individual offers no suggestion as to what these subdivisions, if any, must be, whereas to conceive a segment as a class imposes a definite scheme of subdivision—into subclasses and members. ([17], p. 45).

For example, the set of human beings distinguishes one segment of the total universe, human beings, from all else. If we conceive of human beings as a whole or individual, a whole with parts scattered across space through time, we single out the same segment of the universe that the set of human beings isolates, but without prescribing a scheme for subdividing the segment thus marked off. Conceived as a whole or individual, the segment *human beings* belongs to the same logical type as Socrates; neither is a set but both can be members of sets. By contrast, the set of human beings belongs to a logical type quite different from that to which Socrates or Plato belongs. The set of human beings has members and is not a material object; Socrates and Plato have no members (they aren't sets) but are material objects. Socrates' pug nose is a part of the segment human beings but is neither part nor member nor subset of the set of human beings. Socrates himself is a member, but not a part, of the set of human beings; he is a part, but not a member, of the whole or individual that is the segment *human beings*.

Similarly, we can conceive of Tom and Dick together as a set, the set whose members are Tom and Dick, or as an individual or whole whose person parts are Tom and Dick. We shall let $\{Tom, Dick\}$ be the set of Tom and Dick, and let $+(Tom, Dick)$ be the individual or whole whose person parts are Tom and Dick. We shall speak of $+(Tom, Dick)$ or $+(Tom, Dick, Harry)$ as the *sum* or *sum individual* of Tom and Dick, or of Tom, Dick, and Harry, respectively. The sum of Tom and Dick, i.e., $+(Tom, Dick)$, can be or do many things that a person can be or do. Like a person, $+(Tom, Dick)$ has a definite mass, is colored, and is a spatio-

temporal object (albeit a discontinuous one unless Tom and Dick are Siamese twins). Like a person, $+(Tom, Dick)$ can carry things, write books, grow larger or smaller, and have Harry as a shipmate. Assuming, as we have been, that Tom is not Dick, we note that there are very few things that a person can be or do that $+(Tom, Dick)$ cannot be or do. The few *prima facie* exceptions are such as these: being a person, soloing in an airplane, single-handling a yacht.

The *braces operator* enables us to refer to the set $\{a_1, \dots, a_n\}$ of any finite number of individuals $a_1, \dots, a_n$ that we can designate. Similarly, the *plus-sign operator* enables us to denote the sum individual $+(a_1, \dots, a_n)$ of such individuals. Thanks to the *circumflex operator*, we can refer to the set of all individuals that have the property φ , i.e. to the set $\hat{x}\varphi x$. We need an analogous device to denote the sum of all individuals that have φ . Leonard & Goodman use the set $\hat{x}\varphi x$ to get to the sum of the φ individuals by introducing a *fusion operator* f , where $f(\hat{x}\varphi x)$ is the individual y such that, for any z , y and z have a part in common if and only if z has a part in common with at least one member of $\hat{x}\varphi x$. Since our objective is to put mereological predicate logic to certain logical and linguistic uses, we forego this detour through set theory in favor of a *sum operator*, the *sigma operator*, that achieves the same effect as f . This sum operator is reminiscent of the *iota* or definite description operator. Recall that $(\iota x)\varphi x$ is the one and only φ . Should there be no φ , or more than one φ , " $(\iota x)\varphi x$ " fails to denote, occasioning truth-value gaps.³ Analogously, we let $(\sigma x)\varphi x$ be the sum of all the φ s; if there are no φ s, " $(\sigma x)\varphi x$ " fails to denote, giving rise to truth-value gaps.⁴ Letting " $Ox\bar{x}$ " mean " x overlaps y " (i.e., " x and y have a part in common"), we can define the sigma operator in terms of the *iota operator* thus:⁵

$$(8) (\sigma x)\varphi x =_{df} (\iota x)(\forall y)[Oyx \equiv (\exists z)(\varphi z \ \& \ Oyz)]$$

The circumflex operator renders the braces operator superfluous, for $\{a_1, \dots, a_n\}$ can be explained as $\hat{x}(x = a_1 \vee \dots \vee x = a_n)$. Similarly, when the sigma operator is available, one can dispense with '+' by defining $+(a_1, \dots, a_n)$ as $(\sigma x)(x = a_1 \vee \dots \vee x = a_n)$.

³ There are many ways to close these gaps. We shall adopt Burge's method which consists in treating as false simple sentences containing non-denoting singular terms. For exposition of this method and arguments that favor its use when a semantics for English is at issue, see Burge [3].

⁴ In both appearance and meaning, our sum operator bears a close resemblance to Parsons' substance-abstraction operator. See Parsons [21] or Burge [2].

⁵ Though he introduces no special sum operator, Goodman supplies the *definiens* of our sum operator in [9], p. 52.

III. MULTIGRADE PREDICATES AND PERASAL CONJUNCTION

Leonard & Goodman call predicates that take variably many, as opposed to a fixed number of arguments, *multigrade* predicates.⁶ Predicates that take a fixed number of arguments are characterized as *unigrade*. In their own words ([17], p. 50), when we confront a multigrade predicate "we encounter a relation which is significant in varying degrees; as dyadic, as triadic, tetradic, and so on." The relation "met with," for example, may obtain between two people, among three, among four, and indeed among any number. Such a relation without any fixed degree may be called a "multigrade relation." Thus "a unigrade relation is a relation of any one degree; a multigrade relation is one having at least two different degrees."

For reasons that will become apparent, we will modify slightly Leonard & Goodman's terminology. First, we shall regard all predicates as having a fixed number of argument places. Thus we construe "is a lawyer" as monadic, "is a shipmate of" as dyadic, and so on. (The predicate "are shipmates" will be seen to be a surficial manifestation of an underlying dyadic predicate.) We shall say that a predicate is *multigrade at* an argument place iff it admits certain special terms denoting sum individuals at that argument place. These special terms, which following Lakoff & Peters we shall call NP*, will be explained later; for the moment it will suffice to say that "Tom, Dick, and Harry," as used in sentence (9),

- (9) Tom, Dick, and Harry carried the piano upstairs.

is a paradigm NP*.⁷ Thus sentence (9) shows that the dyadic predicate "carry" is multigrade at its first argument place. Sentences (10) and (11)

- (10) Rev. Brown and Rev. Smith married John and Mary.
(11) Tom and Dick killed Harry.

show that the dyadic predicates "marry" and "kill" are multigrade at their first and second and at their first argument places respectively. (Sentence

(10) envisages a marriage performed jointly by two ministers.)

By a *multigrade predicate* we shall mean a predicate that is multigrade at one or more of its argument places. [As will become evident later, we lean toward the view that nearly all predicates are multigrade predicates that are multigrade at all their argument places.] We maintain that the predicates called *symmetric* by Lakoff & Peters are the very predicates that we have called multigrade (see [15]). Lakoff & Peters call them *symmetric* because any permutation of the conjuncts of their NP* arguments yields a sentence that qualifies as a synonymous paraphrase of the original.

For example, such permutations yield (12)–(14)

- (12) Dick, Harry, and Tom carried the piano upstairs.
(13) Rev. Smith and Rev. Brown married Mary and John.
(14) Dick and Tom killed Harry.

as synonymous paraphrases of (9)–(11) respectively. This symmetry phenomenon was already noted and explained by Leonard & Goodman [17]. It constitutes a salient logical feature that must be accounted for by mereological predicate logic.

Like all predicates, multigrade predicates have extensions. Extensions, moreover, are relative to the domain of discourse, i.e., the set of objects that constitutes the range of the variables of quantification.⁸ For example, the extension of the monadic predicate "is red" will depend on what set is taken as the domain of quantification; if that domain contains nothing that is red, its extension will be the empty set. The first step in interpreting ordinary predicate logic is to specify a set as the domain of quantification. Let's call this set the *atomic domain*. The atomic domain **D** induces a larger domain **D*** which contains not only all the elements of **D** but also all sums of elements of **D**. Where *N* is the cardinality of **D**, the cardinality of **D*** will be $2^N - 1$. That is, corresponding to each *nonempty* subset of **D** there will be an element of **D*** that is the sum of the members of that subset. We shall call the domain **D*** induced by the atomic domain the

⁶ The braces operator of " $\{a_1, \dots, a_n\}$ " and the '+' operator of " $+(a_1, \dots, a_n)$ " are bonafide multigrade operators. These multigrade operators as well as multigrade sentence connectives are studied in Hendry & Massey [12]. Recently, McCawley has argued that English deep structures require multigrade sentence connectives. See McCawley [20], pp. 519 ff.

⁷ Sentence (9) is of course ambiguous. We have in mind the reading that contrasts with the paraphrase "Tom, Dick, and Harry each carried the piano upstairs."

⁸ The extension of a predicate is relative to the domain of discourse in the sense that changes in the domain can affect the extension of an interpreted predicate, a phenomenon dramatized by the so-called Skolem paradox.

domain of individuals. In mereological predicate logic, it is the domain of individuals that serves as the range of the variables of quantification, subsets of this domain serve as extensions of monadic predicates, and so on. The leading idea is that one talks about, i.e., quantifies over, not just atomic individuals (the members of the atomic domain) as in ordinary predicate logic, but all the individuals in the induced domain, whether atomic or not. (We have been assuming that the purposes that led one to select a set as the atomic domain allow one to treat all its members as mutually discrete, i.e., as having no parts in common. When this assumption fails, some modifications of the foregoing account must be made.)

The extension of a monadic predicate, unigra- or multigra- is the set of things (in the domain of individuals) that the predicate is true of. Some of these things may be ordinary or atomic individuals; others may be sum individuals. For example, if we let the atomic domain be the set of human beings, "weighs over 270 pounds" is a monadic predicate that is true of the Pittsburgh Steelers' defensive tackle Joe Greene, true of + (Gerald Ford, Pierre Trudeau), true of + (Golda Meir, Brigitte Bardot, Gloria Steinem), and true of $(\sigma x)(x$ is a first-grade pupil at Sacred Heart School in Pittsburgh); it is false of Gerald Ford, false of + (Brigitte Bardot, Gloria Steinem), and false of $(\sigma x)(x$ is a present Queen of England). So Joe Greene, + (Gerald Ford, Pierre Trudeau), + (Golda Meir, Brigitte Bardot, Gloria Steinem) and $(\sigma x)(x$ is a first grade pupil at Sacred Heart) all belong to the extension of "weighs over 270 pounds." Or, what comes to the same, the following sentences are all true:

- (15) Joe Greene weighs over 270 pounds.
- (16) Gerald Ford and Pierre Trudeau weigh over 270 pounds.
- (17) Golda Meir, Brigitte Bardot, and Gloria Steinem weigh over 270 pounds.
- (18) All the first grade pupils at Sacred Heart School weigh over 270 pounds.

[Note that in (18) we analyzed "all the ϕ " as " $(\sigma x)\phi x$," treating it as a *group quantifier*. The distinction between ordinary distributive quantifiers and group quantifiers will be explained in §6 below.]

A bygone grammar called the subjects "Gerald Ford and Pierre Trudeau" and "Golda Meir,

Brigitte Bardot, and Gloria Steinem" of (16) and (17) *compound nouns*; nowadays they are called *coordinate conjoined* NPs (noun phrases) or for short NPⁿs. The NPⁿs of (16) and (17) are both NP^{*}s. Hitherto, we offered only a rough and ready characterization of NP^{*}s; it is time to say precisely what they are. We define an NP^{*} to be an NPⁿ that denotes the sum of the individuals denoted by its coordinated NPs. For example, the NPⁿ of (16) is an NP^{*} because it denotes + (Ford, Trudeau), i.e., it denotes the sum of what "Gerald Ford" and "Pierre Trudeau" denote.

What was just said needs qualification. Sentence (16) is ambiguous; it can mean either (19) or (20).

- (19) + (Ford, Trudeau) weighs over 270 pounds.
- (20) Ford weighs over 270 pounds, and Trudeau weighs over 270 pounds.

The NPⁿ of (16) is an NP^{*} only when (16) is read as (19). Departing slightly from present linguistic usage, we shall refer to NPⁿs that are NP^{*}s as *phrasal conjunctions*. And we shall use "phrasal conjunction" ambiguously, sometimes meaning an NP^{*}, sometimes a sentence containing an NP^{*}. Thus (16) and (17) can be read either as phrasal or as non-phrasal conjunctions. Although (18) contains no NPⁿ, it is similarly ambiguous between a group-quantifier reading (21)

- (21) $(\sigma x)(x$ is a first-grade pupil at Sacred Heart) weighs over 270 pounds.

and a distributive-quantifier reading (22).

- (22) $(\forall x)(x$ is a first-grade pupil at Sacred Heart $\supset x$ weighs over 270 pounds)

Of course, neither (17) nor (18) ordinarily strikes us as ambiguous, because extra-linguistic information causes us unreflectively to dismiss their respective non-phrasal and distributive readings as non-intended.

So far as Tarskian truth definitions go, multigra- and unigra- predicates are handled alike. For example, the satisfaction clause for "weighs over 270 pounds" might be stated thus: $(\forall x)(x$ satisfies "weighs over 270 pounds" if and only if x weighs over 270 pounds). The metalinguistic quantifier must range over all the individuals in the domain of individuals of the object language, but that holds for truth definitions generally. The only difference is that the relevant object-language domain must be the induced domain.⁹

⁹ Burge makes the same point about satisfaction clauses for mass terms which he treats as true of, or false of, objects in what we have called the induced domain. See [2], p. 274 and p. 282.

Like ungrade predicates, multigrade predicates come in all degrees; some are monadic, some dyadic, some triadic, and so on. For example, "marry" is a dyadic multigrade predicate; its extension is a set of ordered pairs the elements of which can be ordinary individuals or sum individuals. For example, given that (10) is true, the ordered pair $\langle +(\text{Rev. Brown, Rev. Smith}), +(\text{John, Mary}) \rangle$ belongs to the extension of "marry."¹⁰

The order of the singular terms inside braces is immaterial to what set the whole expression denotes; for example, $\{\text{Tom, Dick, Harry}\} = \{\text{Harry, Tom, Dick}\}$. The order of the arguments of '+' is similarly immaterial to what individual the entire expression denotes; for example, $+(\text{Tom, Dick, Harry}) = +(\text{Harry, Tom, Dick})$. It follows, then, that phrasal conjunctions will exhibit the symmetry phenomenon mentioned in §3 above. Moreover since (25).

- (25) $(\forall x_1) \dots (\forall x_n) [+ (x_1, \dots, x_n) = + (x'_1, \dots, x'_n)]$, where $(x'_1, \dots, x'_n)$ is a permutation of $(x_1, \dots, x_n)$.

is a theorem schema of mereological predicate logic, the latter can certify as valid any intuitively valid argument whose validity, like that of (A1), turns on the symmetry of phrasal conjunctions. In addition to (25), only rules governing identity are needed to validate such arguments. Identity itself is definable within mereological predicate logic in myriad ways; for example, one can use the predicate "is a part of," or for short, '<', to explain " $x=y$ " as " $(\forall z)(z < x \equiv z < y)$."

Not everything that glitters is gold, nor is every NP^a an NP*. The NP^a "Tom, Dick, and Harry" of (26)

- (26) Tom, Dick, and Harry are linguists.

is not an NP*, i.e., not a phrasal conjunction. Sentence (26) affirms "linguisthood" of Tom, of Dick, and of Harry, but not of $+(\text{Tom, Dick, Harry})$. Put otherwise, (26) has three logical subjects; were (26) a phrasal conjunction, it would have just one such subject. The logical form or deep structure of (26), then, is (27) rather than (28).¹¹

- (27) Linguist(Tom) & linguist(Dick) & linguist(Harry)

- (28) Linguist $(+(\text{Tom, Dick, Harry}))$

Lakoff & Peters apply the term *sentential conjunction* to those surficial NP^as that result when a so-called *conjunction reduction transformation* transforms an underlying conjunction of sentences such as (27) into a superficially simplex sentence like (26).¹² The question arises, then, whether every NP^a is either a phrasal or a sentential conjunction. We argue below that the NP^as of (29)–(32).

- (29) Tom, Dick, and Harry are similar.
(30) Tom, Dick, and Harry like each other.
(31) Tom, Dick, and Harry punched one another.
(32) Tom, Dick, and Harry are partners.

are neither phrasal nor sentential conjunctions. Sentences (29)–(31) are akin to sentential conjunctions; (32) is a mongrel sentence that has both sentential and phrasal aspects.

IV. THE UNIGRADE BIAS OF PHILOSOPHY

A leading theme of the Aristotelian logical tradition attributed subject-predicate form to all simple (atomic) propositions. Ontologically, this meant that the only real or genuine properties were monadic properties; polyadic properties (relations) were pseudo-properties, illusions to be explained away. So an analysis of a relation was deemed successful within this tradition only if it reduced the relation to monadic properties. For example, a successful analysis of the relation *outrank* among face cards might reduce it to monadic properties thus: x outranks $y =_{df}$ (Queen x & Jack y) $\vee$ (King x & Jack y) $\vee$ (Ace x & Jack y) $\vee$ (King x & Queen y) $\vee$ (Ace x & Queen y) $\vee$ (Ace x & King y). This bias against relations had a stranglehold on logic and philosophy for two millenia. It had, of course, its successes the greatest of which was Leibniz's reduction of the relation of *identity* (and *diversity*) to monadic properties: x is identical to $y =_{df}$ $(\varphi)(\varphi x \equiv \varphi y)$. (The Kneales artfully chronicle how the traditional depreciation of relations stunted much of Leibniz's work and gave it its peculiar cast. See [13], pp. 320 ff.) It was not until the 19th century that the irreducibility and im-

¹⁰ Throughout this paper we ignore matters of tense.

¹¹ The only rule of insertion of grouping indicators that we follow is to put them wherever they contribute to conceptual perspicuity.

¹² Linguists apply "conjoined" and its cognates to every manner of linking like constituents in a coordinate way. Thus "Tom, Dick, or Harry" and "neither Tom, Dick, nor Harry" are no less NP^as than is "Tom, Dick, and Harry." For the most part, we follow their practice.

portance of relations were fully appreciated and a logical theory and notation developed that treated polyadic and monadic properties evenhandedly. In the absence of these conceptual and notational developments, the astonishing achievements of the past 100 years in logic and mathematics would have been impossible.

It is universally recognized that the conceptual and notational resources of the logic of Frege and Whitehead-Russell are adequate for classical mathematics. That is to say, this logic furnishes logical forms (deep structures) for the propositions of classical mathematics that explain and justify its axiomatic architectonic; every recognized inference is shown to be valid; every established mathematical theorem, to be a logical consequence of the relevant set of postulates. (Logicians claim even greater achievements for this logic, viz. that every mathematical concept can be defined by means of logical concepts in such a way that the theorems of classical mathematics become logical truths. Unlike its axiomatic success mentioned above, this putative reductive triumph of modern logic has often and seriously been challenged.) The logic of Frege and Whitehead-Russell makes no provision for multigrade predicates, so *a fortiori* classical mathematics can get along without them. Within arithmetic, their absence is not felt because arithmetical discourse, along with set discourse, is unique in that none of its predicates present even a *prima facie* appearance of multigradity. One might object that the equality predicate of (33) and (34)

(33) 2 and 2 equal 4.

(34) 2 and 2 equals 4.

is *prima facie* multigrade since "2 and 2" appears to be a phrasal conjunction. But if "2 and 2" were a phrasal conjunction, (35)

(35) 2 and 2 equals 2.

would be a mathematical truism for the reason that $+(2, 2) = 2$. (Remember that '+' is an ontological sum operator, not the addition operator. To prevent confusion, we will write the latter as '⊕'.) Reporting on her linguistic field work, Gleitman ([7], p. 291) notes that nearly all informants confessed that they had often been puzzled by sentences like (33) and (34), not knowing whether to employ the singular or the plural form of the verb. This suggests that English speakers perceive, however dimly, that the "and" of (33)–(35) is neither the "and" of sentential conjunction nor the "and"

of phrasal conjunction. Mathematicians, of course, treat this "and" as denoting a dyadic operation on numbers, thereby assigning to both (33) and (34) the deep structure (36)

(36) $(2 \oplus 2) = 4$.

Unlike arithmetic, geometry harbors *prima facie* multigrade predication. But the multigrade predicates of geometry are *systematically* explained away by an ontological reduction that construes all geometrical objects as point sets. (The points themselves are often further reduced to *n*-tuples of real numbers or the like. This second reduction is sufficient to eliminate multigrade predicates, but is not necessary.) Consequently, multigrade geometrical predicates can be systematically replaced by the unigrade predicates of the reductive theory of point sets. For example, the geometrical predicate "make up" of (37).

(37) (Line segments) AB, BC, and AC make up (triangle) ABC.

is multigrade, since "AB, BC, and AC" is a phrasal conjunction that denotes $+(AB, BC, AC)$. But by taking line segments and triangles to be sets of points, one can represent the deep structure of (37) by (38) in lieu of (39).

(38) $(AB \cup BC \cup AC) = ABC$

(39) $+(AB, BC, AC) = ABC$

The advantage of (38) over (39) is this: (38) can be handled by the logic of Frege and Whitehead-Russell that makes no provisions for multigradity whereas (39) demands just such provisions.

The overt unigradity of arithmetical and set-theoretic predicates, together with the covert (reductive) unigradity of geometrical predicates, explain the success of predicate logic as a theory of the deep structure of mathematical propositions as well as the nearly universal failure of logicians to perceive, much less appreciate, the multigrade character of non-mathematical English predicates. And since philosophers naturally take their formal cues from logicians, one can predict that philosophy will be largely or wholly insensitive to multigrade and phrasally conjunctive phenomena. Consider, for example, epistemology. Epistemologists proceed as though so-called *mental verbs* like "know" and "believe" were literally predicates solely of persons and only derivatively predicates of person-sums or other individuals. That is, if they notice it at all, epistemologists regard the multigrade use of mental predicates as parasitic upon their ostensibly

unigrade) application to persons in much the same way that the Aristotelian tradition treated relations as parasitic on properties. No wonder, then, that such elements as awareness or consciousness often find their way into analyses of belief and knowledge. Someone who treated evenhandedly knowledge attributions like (40) and (41)

(40) Tom knows the answer.

(41) Tom and Dick know the answer, although neither Tom nor Dick knows it.

would likely assign a lesser role to consciousness in a theory of knowledge than would a unigrade-biased philosopher. Perhaps the multigrade use of "know" can be reduced to person predications of "know," but that is a question to be decided by evidence and argument, not by a biased methodology.

The primacy of person predication, which is but one manifestation of the unigrade bias of philosophy, has also pervaded ethics. Only recently have some moral philosophers, notably Virginia Held, begun to see that such ethical notions as guilt and responsibility are *prima facie* neutral between person and person-sum predication.¹³ Equipped with a logico-semantic theory of phrasal conjunction and group quantification, together with an appropriate symbolic notation, these perspicacious moral philosophers should be able to articulate and develop their novel insights with a facility that would otherwise be impossible.

Though many more could be given, we will take action theory as our last philosophical illustration of the unigrade bias. Goldman's *A Theory of Human Action* purports to provide a general theory of human action, but in fact offers only a theory of the actions of individual human beings. Nothing is said about such actions as the action of Tom, Dick, and Harry described by sentence (9), nor is such omission noted. On the next to last page of the book Goldman remarks that he has "focused on the acts of single agents, rather than the relationship between acts of several agents" and commends his theory for being extendable to "the interactions of two or more agents" ([8], p. 225). But (9) *prima facie* describes an act of a single agent, namely +(Tom, Dick, Harry), not the interaction of three agents. Perhaps such actions can be reduced to the actions of individual persons, but again, that is a matter to be settled by evidence and argument, not by tacit appeal to the reductive thesis ingredient in the unigrade bias.

Nor is it evident that the actions of +(Tom, Dick, Harry) are identical with or reducible to *group actions*, i.e., the actions of *collectives* such as families or nations. In addition to the actions and interactions of persons, Goldman recognizes that there may also be bonafide actions of groups or collectives the explanation of which might require a model of action different from his own ([8], p. 226). What is not taken into account in his otherwise impressive and comprehensive book are the actions of such agents as +(Tom, Dick, Harry).

Along with Leonard and Goodman, Wilfrid Sellars is the philosopher who has most clearly recognized the importance to metaphysics of an account of NP's and multigrade predicates. Sellars calls the apparent denotata of non-overtly-sentential NP's *conjunctive individuals*. He declares that "unlikely though it may seem [conjunctive individuals] pose one of the central problems in the metaphysics of persons and sentient beings generally. For to say that conjunctive individuals are reducible is to say that statements about them can be paraphrased in ways which refer only to their 'constituents' (i.e., conjuncts)" ([25], p. 199). Sellars' conjunctive individuals outstrip somewhat our sum individuals, for Sellars associates conjunctive individuals not only with NP's but also with such non-phrasal NP's as those in (29)–(32). Nevertheless, we do Sellars no disservice by reformulating the aforementioned metaphysical problem as a problem about *sum individuals*, in particular about sums of persons. The problem is two-fold. First, there is the *emergentist* puzzle. Do person sums have properties or perform actions irreducible to the properties or actions of persons? Or equivalently, must a comprehensive inventory of the Manifest Image include entries for sums of persons? The second puzzle is *reductive*. The Scientific Image seems to leave no room for persons, save as sums of elementary particles. Does the substantial unity of persons, the Manifest Image's paradigm substances, dissolve, then, into the accidental unity of elementary particle sums?

V. GRAMMARS OF PHRASAL CONJUNCTION

(Coordinate) compound constituents other than compound sentences have been largely neglected by logicians and philosophers, but they have lately commanded much attention from generative grammarians. [Compound constituents are known

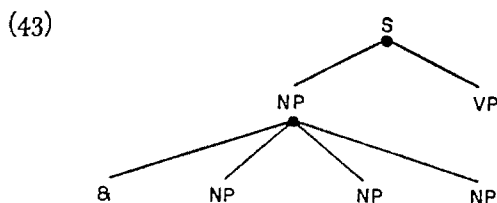
¹³ See Held [11]. Held's *random collections of individuals* are sums of persons.

as *coordinate conjoined structures*; they include compound noun phrases alias NP's, compound verb phrases alias VP's, and compound sentences alias S's.] For want of a guiding semantics, all the grammars of compound constituents known to us are radically defective. Justification and documentation of this sweeping condemnation must be deferred to another time and place. Here we can describe only the character of these grammars and offer general criticisms of their several approaches to compound constituents. To keep matters tractable, we limit discussion to NP's.

Lakoff & Peters' grammar (See [15]) is typical of those theories that lump together as "phrasal" all NP's that are not sentential, i.e., that do not arise from underlying sentential conjunction via some species of conjunction reduction transformation. The core idea is to introduce "phrasal" NP's directly into deep structure by means of phrase-structure rules like (42).

(42) $NP \rightarrow NP^n$

The "phrasal" NP's generated by (42) contrast with sentential NP's which do not appear at all in deep structure; sentential NP's are derived constituents of intermediate or surface structure, generated from deep-structure S's via reduction transformations. Thus Lakoff & Peters' grammar assigns deep structures of the form (43)



to all of (44)–(47)

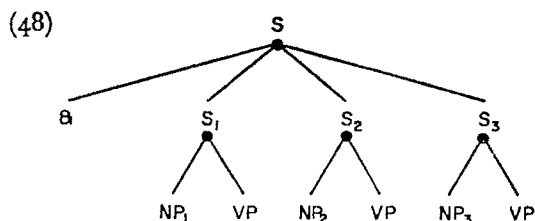
(44) Tom, Dick, and Harry carried the piano upstairs.

(45) Tom, Dick, and Harry are similar.

(46) Tom, Dick, and Harry conferred.

(47) Tom, Dick, and Harry are partners.

while assigning deep structures of the form (48)



to sentences (49) and (50)

(49) Tom, Dick, and Harry are lawyers.

(50) Tom, Dick, and Harry carried the piano upstairs.

when (50) is used to report three feats of strength. To generate (49), of course, the grammar must also have a phrase-structure rule like (51).

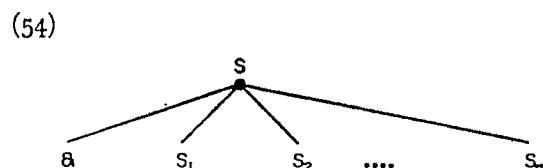
(51) $S \rightarrow S^n$

A grammar of a language L is intended to generate all and only the well-formed expressions of L . *Inter alia*, a grammar is supposed to generate the well-formed expressions of L in a way that will account for whatever deductive logical relations obtain among them. Such an account normally has the character of a semantical theory of the *deep-structure* schemata of L , these schemata being the deep structures minus their topic-specific lexical items. [The *topic-specific* lexical items may be thought of as the *words* that are introduced by *terminal* phrase-structure rules like (52) and (53):

(52) $V \rightarrow \text{run, play, etc.}$

(53) $N \rightarrow \text{girl, boy, ball, etc.}$

Lexical items introduced by *non-terminal* phrase-structure rules like (42) and (51) are not topic specific, and so may occur as parts of deep-structure schemata. A phrase structure rule is *terminal* just in case its application yields only nodes to which phrase-structure rules are inapplicable.] For example, (54)



is a deep-structure schema. Corresponding to (54), Lakoff & Peters' grammar might contain a semantical rule like (55):

(55) $\&(S_1, \dots S_n)$ is true if $S_1, \dots S_n$ are all true; otherwise $\&(S_1, \dots S_n)$ is false.

Semantical rule (55) accounts for the symmetry phenomenon of *sentential* conjunction noted in §3 above, i.e., it accounts for the validity of such inferences as the one from (49) to (56).

(56) Dick, Harry, and Tom are lawyers.

It also accounts for the following *contraction phenomenon*, as exhibited by *sentential* conjunctions:

the inference from $\varphi(NP^{n+1})$ to $\varphi(NP^n)$ is valid where $n \geq 1$. [When $n=1$, NP^n is the same as NP .] That is, rule (55) explains why inferences like the one from (56) to (57)

(57) Dick and Harry are lawyers.

are valid. In addition, rule (55) explains why sentential NP^n s expand, i.e., it accounts for the validity of the inference to $\varphi(a_1, \dots, a_n \text{ and } a_{n+1})$ from the two premisses $\varphi(a_1, \dots, a_n)$ and $\varphi(a_{n+1})$. In particular, rule (55) accounts for the validity of the inference to (56) from the premisses (57) and (58).

(58) Tom is a lawyer.

Since symmetry, contraction, and expansion are the only relevant logical phenomena exhibited by sentential NP^n s, Lakoff & Peters' grammar furnishes an account of the deep structure of sentential conjunction that is adequate from a logical point of view.

Unfortunately, the same cannot be said of Lakoff & Peters' account of "phrasal" conjunction. By lumping together deep-structure schemata whose logical behavior is very different, Lakoff & Peters' phrase-structure rule (42) makes it impossible to formulate a semantical rule that will account for the deductive relationships of "phrasal" conjunctions in the way that rule (55) accounts for the logic of sentential NP^n s. At first blush it might seem that the "phrasal" conjunctions (44)–(47) are alike with respect to their deductive behavior; all seem to exhibit symmetry and contraction, but not expansion. A closer look dispels this illusion. Note, first, that (44) does not contract. "Phrasally" interpreted (44) does not imply (59).

(59) Tom and Dick carried the piano upstairs.

[The failure of this contractive implication to obtain becomes transparent when the VP "weigh 600 pounds" is substituted for the VP in (44) and (59).] To represent the deep structure of sentences like (44), Lakoff & Peters need a phrase-structure rule like (60):

(60) $NP \rightarrow + (NP_1, \dots, NP_n)$

Corresponding to (60), a semantical rule like (61)

(61) The NP " $+ (NP_1, \dots, NP_n)$ " denotes the sum of the individuals denoted by $NP_1, \dots, NP_n$.

can be formulated that accounts completely for the logic of the phrasal conjunctions introduced into

deep structure by (60). These NP^n s are precisely those that were called NP^* s in §3.

Secondly, notice that (45)–(47) do contract so long as the contracted NP^n contains at least two NPs, i.e., so long as $n > 1$. For example, (45) and (47) respectively imply (62) and (63).

(62) Tom and Dick are similar.

(63) Tom and Dick are partners.

but they do not imply the ill-formed sentences (64) and (65).

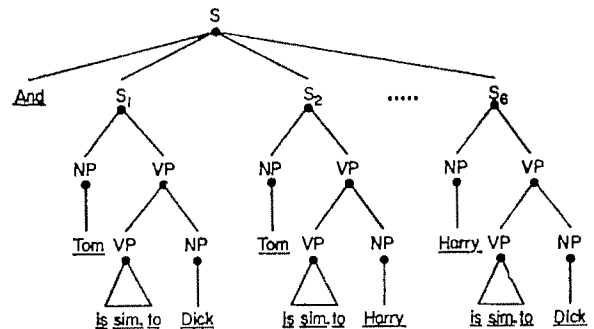
(64) *Tom is similar.

(65) *Tom is partners.

The foregoing observation suffices to demarcate (45)–(47) from (44) so far as their logical behavior is concerned. The question remains, of course, whether (45)–(47) form a natural class of NP^n s, or whether they represent from a logical point of view quite different types of NP^n s, as different from one another as sentential NP^n s are different from phrasal NP^* s.

Gleitman [7] has proposed a grammar of NP^n s that, unlike Lakoff & Peters' grammar, introduces no NP^n s into deep structure. NP^n s are one and all derived constituents generated by transformations. According to Gleitman's grammar, (45) has the deep structure (66).

(66)



Conjunction reduction transformations, followed by an *each other* transformation, carry (56) into (67).

(67) Tom, Dick, and Harry are similar to each other.

which, upon application of an optional deletion transformation, becomes the surface-structure sentence (45). Gleitman's general approach to the deep structure of sentences like (45) is very attractive. In tandem with semantical rule (55), it accounts for the symmetry and contraction phe-

nomena that such sentences exhibit. It also explains why contracted implicates must have at least two NPs in their NPⁿs; the underlying predicate is transitive (dyadic) and so calls for both a subject NP and an object NP.

Unquestionably, one reading of (45) is the *each other* reading captured by Gleitman's grammar. Let NPⁿR be schematic for sentences like (45), where R is a dyadic (transitive) deep-structure predicate masquerading as a surface monadic (intransitive) predicate. Then Gleitman's grammar assigns to NPⁿR the deep structure (68):

$$(68) \quad \& (R NP_i NP_j)^k, \text{ for all } i \neq j, \quad 1 \leq i \leq n, \\ 1 \leq j \leq n, \quad k = n^2 - n.$$

But (45) can also receive a weaker *one another* reading paraphrased by (69).

$$(69) \quad \text{Tom is similar to Dick or Harry, Dick is similar to Tom or Harry, and Harry is similar to Dick or Tom.}$$

The *one another* reading of a surface sentence NPⁿR answers to deep structure (70).

$$(70) \quad \& (\vee (R NP_i NP_j)^{n-1})^n, \text{ for all } i \neq j, \\ 1 \leq i \leq n, \quad 1 \leq j \leq n.$$

It can be readily shown that (70) and (55), together with semantical rule (71),

$$(71) \quad \vee (S_1, \dots, S_n) \text{ is true just in case at least one of } S_1, \dots, S_n \text{ is true.}$$

captures the logical properties of NPⁿR sentences with *one another* readings. Semantical rule (71) itself corresponds to the phrase-structure rule (72).

$$(72) \quad S \rightarrow \vee (S_1, \dots, S_n)$$

It would be an easy task, then, to extend Gleitman's grammar to handle *one another* sentences.

Sentences (46) and (47), however, resist Gleitman's approach. Though (46) implies both (73) and (74)

$$(73) \quad \text{Tom, Dick, and Harry conferred with each other.}$$

$$(74) \quad \text{Tom, Dick, and Harry conferred with one another.}$$

neither (73) nor (74) implies (46). Similarly, (47) implies both (75) and (76)

$$(75) \quad \text{Tom, Dick, and Harry are partners of each other.}$$

$$(76) \quad \text{Tom, Dick, and Harry are partners of one another.}$$

neither of which implies (47). (For (47) to be true, there must be some firm to which Tom, Dick and Harry all belong; for (75) and (76) to be true, it suffices that there be three firms such that Tom and Dick belong to the first, Tom and Harry to the second, and Dick and Harry to the third.) One of Leonard & Goodman's most important contributions was an ingenious analysis of the logical form of sentences like (46) and (47). In giving the deep structure of such sentences, we depart slightly from their analysis. Part of the motive for our departure is to ensure that these sentences contract; for reasons connected with their constructionalist enterprise, Leonard & Goodman did not want these structures to exhibit the (modified) contraction phenomenon.

Because the cumbersome notation of grammarians obscures somewhat the analysis about to be presented, we will revert to the notation of mereological predicate logic. When rewritten in this notation, Gleitman's proposed deep structure (68) of surface sentences " $\Omega(a_1, \dots, \text{and } a_n)$ " with *each other* readings becomes (77):

$$(77) \quad \& (\Omega a_i a_j) \text{ for all } a_i, a_j, i \neq j, 1 \leq i \leq n, 1 \leq j \leq n.$$

For the sake of future perspicuity, we unpack (77) to get (78).

$$(78) \quad [\Omega a_1 a_2 \& \dots \& \Omega a_1 a_n] \& [\Omega a_2 a_1 \& \Omega a_2 a_3 \\ \& \dots \& \Omega a_2 a_n] \& \dots \& [\Omega a_n a_1 \& \dots \& \\ \Omega a_n a_{n-1}]$$

Whatever deep structure we may propose for sentences like (46) and (47), then, it must imply (78) as well as the *one another* reading of " $\Omega(a_1, \dots, \text{and } a_n)$ " whose deep structure is given by (79).

$$(79) \quad [\Omega a_1 a_2 \vee \dots \vee \Omega a_1 a_n] \& [\Omega a_2 a_1 \vee \Omega a_2 a_3 \\ \vee \dots \vee \Omega a_2 a_n] \& \dots \& [\Omega a_n a_1 \vee \dots \vee \\ \Omega a_n a_{n-1}]$$

[Since (78) itself implies (79), whatever implies (78) will also imply (79).] Perspicuity will also be served by two preliminary abbreviative conventions. We shall let " $\Omega a(b_1 \& \dots \& b_k)$ " abbreviate the conjunction " $\Omega a b_1 \& \dots \& \Omega a b_k$," and shall let " $b_1? + \dots + b_k?$ " abbreviate the lengthy "conjunction" whose $2^k - 1$ "conjuncts" are all the terms of the form " $+(b_1, \dots, \beta_m)$," $1 \leq m \leq k$, where $\langle \beta_1, \dots, \beta_m \rangle$ is any subsequence of $\langle b_1, \dots, b_n \rangle$; for example " $+(b_1)$," " $+(b_7)$," " $+(b_2 b_5)$," " $+(b_1 b_3 b_6)$," and " $+(b_1, \dots, b_8)$ "

are all "conjuncts" of " $b_1? + \dots + b_n?$ ". By virtue of these abbreviative conventions, then, (80) abbreviates (81).

$$(80) \quad \Omega a_1(a_2? + a_3?) \& \Omega a_2(a_1? + a_3?) \& \Omega a_3(a_1? + a_2?)$$

$$(81) \quad [\Omega a_1 a_2 \& \Omega a_1 a_3 \& \Omega a_1 + (a_2, a_3)] \& [\Omega a_2 a_1 \& \Omega a_2 a_3 \& \Omega a_2 + (a_1, a_3)] \& [\Omega a_3 a_2 \& \Omega a_3 + (a_1 a_2)]$$

[To make (81) more readable, we have replaced " $+(a_1)$ " by its equivalent " a_1 ".] Now, where " $\Omega(a_1, \dots, \text{and } a_n)$ " is schematic for surface sentences like (46) and (47), its deep structure is given by (82).

$$(82) \quad \Omega a_1(a_2? + \dots + a_n?) \& \Omega a_2(a_1? + a_3? + \dots + a_n?) \& \dots \& \Omega a_n(a_1? + \dots + a_{n-1}?)$$

Intuitively, (82) means that each of $a_1, \dots, a_n$ stands in the relation Ω to the sum of any of the others. Notice that (82) implies Gleitman's *each other* structure (78), but not conversely. Moreover, the parallel between (82) and (78) is obvious; (78) is what (82) becomes when complex singular terms are deleted. Gleitman's analysis of sentences like (46) and (47) fails for want of complex singular terms that denote sum individuals, a failure to be expected in any grammatical theory that incorporates ordinary predicate logic as the logic of its deep structures.

Because (82) exhibits both sentential and phrasal aspects, let's call (the NP's of) sentences like (46) and (47) *mongrel conjunctions*. Our theory of mongrel conjunction accounts in a natural way for many logico-grammatical phenomena, viz: (a) *symmetry*, e.g., the logical equivalence of "Tom, Dick, and Harry are partners" with "Dick, Harry and Tom are partners"; (b) *modified contraction*, i.e., that " $\Omega(a_1, \dots, \text{and } a_{n+1})$ " implies " $\Omega(a_1, \dots, \text{and } a_n)$ " for $n \geq 2$; (c) the *ungrammaticalness* of " Ωa_n ", e.g., the fact that "Tom is partners" is ungrammatical; and (d) the fact that mongrel conjunctions *imply* their *each other* and their *one another* readings, but *not conversely*. A special case of (d) is the non-equivalence of " $\Omega(a_1 \text{ and } a_2)$ " and " a_1 is an Ω of a_2 ." We mention this special case only because Lakoff & Peters' grammar predicts the equivalence of " $\Omega(a_1 \text{ and } a_2)$ " with " a_1 is an Ω of a_2 ," the latter coming by a movement transformation from the former. But this prediction is clearly false: " a is a

brother of b " is hardly equivalent to " a and b are brothers" no matter what Lakoff and Peters' grammar may predict (or Dougherty may think).¹⁴ For Lakoff & Peters, "are partners" and its ilk are *anomalous* deep-structure intransitive VPs that do not take singular subjects; for us they are surficial traces of standard deep-structure transitive VPs that require both a subject and an object, their apparent intransitivity and need for non-singular subjects being merely surface illusions. In addition to (a)-(d) above, perhaps the principal virtue of our theory is that it relates a mongrel conjunction like (47) to the corresponding *group quantification* "All the men are partners" in much the same way that standard logic relates a sentential conjunction like (49) to the corresponding *distributive quantification* "All the men are lawyers." We defer treatment of group quantification, however, to the next section.

Our proclamation that sentences like (46) and (47) have a deep structure patterned after (82), and similar proclamations about other kinds of sentences, have been cavalier in two respects. First, in lieu of clearly and explicitly specifying which sentences have deep structures of the form (82), we have been content to describe them as *sentences like* (46) and (47). Were we presenting a developed grammar of [a fragment of] English, this practice would be reprehensible. But what in fact we purport to offer is a logical analysis, together with a semantical theory, of several kinds of English constructions. And we take advantage of the license logicians have always allowed themselves in such undertakings, viz. to characterize the kind of sentence to which the analysis applies by examples and analogies. Second, we have characterized (82) and other formulas as *deep structures* apart from a system of rules (transformations) that relate them to surface sentences. This practice is a matter of faith, as the first was a matter of license. The *generative semantics hypothesis*, associated with Lakoff, McCawley, Postal, and Ross, identifies deep structure with logical form.¹⁵ We believe that the generative semantics hypothesis, in the slightly more sophisticated form advanced by Bartsch & Venne-mann, is true (see [1], pp. 10 ff.). So having provided the requisite logical forms, i.e., an interpreted system of formulas that captures the relevant implication relations, we anticipate their incorporation by linguists within a grammar of the relevant

¹⁴ See Lakoff & Peters [15], p. 122 f., and Dougherty [6], p. 332.

¹⁵ See, for example, the exposition of the generative semantics hypothesis in Lakoff [14], pp. 588 ff.

fragment of English. Indeed, the absence of such forms to guide the development of their grammars seems to us causally responsible for the serious deficiencies of Lakoff & Peters', Gleitman's, and Dougherty's (see below) grammars.

Notice that the extensions of Ω -predicates are not *symmetric* in the sense that an ordered pair $\langle b, c \rangle$ belongs to the extension of Ω if and only if $\langle c, b \rangle$ also belongs to its extension. In this respect our analysis differs from that of Leonard & Goodman's which invariably effects symmetric extensions. We can find no hard evidence that favors either analysis relative to symmetry of extensions. In the absence of such evidence, simplicity favors our analysis which not only furnishes simpler deep structures for mongrel conjunctions but also holds down the size of the extensions of Ω -predicates as much as possible.

One must not confuse non-symmetry of extensions with a prohibition against sum individuals as *first* elements of the ordered pairs in the extension of an Ω -predicate. Such a prohibition amounts to a diluted version of the unigrade bias, so it is important to see that our theory legislates no such thing. To see this, consider a simple universe U whose atomic domain is the set {Tom, Dick, Mary, Sue}, or, with obvious abbreviations, the set $\{t, d, m, s\}$. The induced domain of U , the set of individuals over which object-language quantifiers range, is then the set $\{t, d, m, s, +(t, d), +(t, m), +(t, s), +(d, m), +(d, s), +(m, s), +(t, d, m), +(t, d, s), +(t, m, s), +(d, m, s), +(t, d, m, s)\}$. Now suppose that Tom, Dick, and + (Mary, Sue) conferred, i.e., suppose that (83)

(83) Tom, Dick, and Mary and Sue conferred.

is true on the reading that takes "Mary and Sue" to denote $+(m, s)$, and suppose that no conferring relations obtain other than those implied by (83). Then the extension of "conferred with" is the set $\{\langle t, d \rangle, \langle t, +(m, s) \rangle, \langle d, t \rangle, \langle d, +(m, s) \rangle, \langle +(m, s), t \rangle, \langle +(m, s), d \rangle, \langle t, +(d, m, s) \rangle, \langle d, +(t, m, s) \rangle, \langle +(m, s), +(t, d) \rangle\}$.

A forceful objection can be brought against our theory of mongrel conjunction, viz. it seems to make it impossible for a person to have exactly *two* shipmates; if Tom, Dick, and Harry are shipmates Tom stands in the shipmate relation to *three* individuals, Dick, Harry, and +(Dick, Harry).¹⁶ To still the anxieties of those who deem it perverse to

countenance person sums among the relata of the *shipmate* relation, Leonard & Goodman ([17], p. 52 f.) suggested that one introduce a technical conjugate "shipmate affiliates" to be treated as an Ω -predicate; then shipmates could be explained as *persons* who are shipmate affiliates. Relative to this maneuver, Tom has three shipmate affiliates but only two shipmates. We think, however, that the maneuver concedes too much to the unigrade bias. There are two other ways to counter the objection. First, one could understand the question "How many shipmates does Tom have?" to ask for the number of *atomic* individuals, persons in the present instance, that are related to Tom by the shipmate relation. Or, second, one could take the question to ask for the number of individuals, atomic or otherwise, that stand in the *shipmate* relation to Tom. Either way, the correct answer *two* is obtained. Our second counter is not available to Leonard & Goodman because the extensions of their Ω -predicates are symmetric. Unlike Leonard & Goodman's, our analysis of Ω -predicates favors their first argument places in a way congenial to the unigrade-minded philosopher. That much due, but no more, are we willing to give to the devil.

Those familiar with Goodman's analysis of similarity classes might deem it strange that we have adopted Gleitman's *each other* and *one another* readings of (62).¹⁷ McCawley ([20], p. 523), for example, rejects Gleitman's analysis on the grounds that Goodman has shown that "the proposition that any two members of a group are similar does not imply that the members of the group are similar since a member of the group might be similar to each other member in a different respect, and there thus might be no similarity shared by all the members, e.g., Tom and Dick might be similar by virtue of looking like Humphrey Bogart, Dick and Harry might be similar by virtue of being confirmed wife-beaters, and Harry and Tom might be similar by virtue of being connoisseurs of pornography, but there might be nothing significant which all three have in common." But Lakoff & Peters have circumscribed a class of predicates having the feature they call *Special* that have the following property: they "require special interpretation, which must be supplied from extralinguistic context or an additional restrictive phrase" ([15], p. 127). The predicate "similar" belongs to this class and so requires special inter-

¹⁶ We owe this clever way of framing the objection to Professor Andrew Oldenquist.

¹⁷ See Goodman [9], pp. 145 ff. and Leonard & Goodman [17], pp. 50 ff.

pretation. What this means is that sentences like (62) cannot stand unaided on their own semantical feet; the respect in which Tom, Dick, and Harry are similar must either be supplied by the extralinguistic context or be made explicit by a restrictive phrase. When this need for special interpretation is recognized, the objection advanced by McCawley is seen to miss its mark. Those unconvinced may treat "similar" as having an Ω -predicate reading over and above its *each other* and *one another* readings.

The last grammar of NP's that we will consider is Dougherty's. Though more fully articulated than Gleitman's or Lakoff & Peters' grammars, Dougherty's grammar (see [5] and [6]) strikes us as semantically the least promising. Gleitman treats all NP's as inherently sentential, and so cannot handle phrasal or mongrel conjunctions. Lakoff & Peters recognize both sentential and "phrasal" NP's, but lump together as "phrasal" all NP's that are not simple sentential conjunctions; i.e., phrasal, mongrel, *each other*, and *one another* conjunctions are all treated as "phrasal" so that the obvious semantical discriminations among them cannot be made. Dougherty, on the other hand, eschews sentential NP's altogether; in his grammar, all NP's and VP's are treated as "phrasal" in deep structure,¹⁸ i.e., all are introduced via phrase-structure rules like (84).

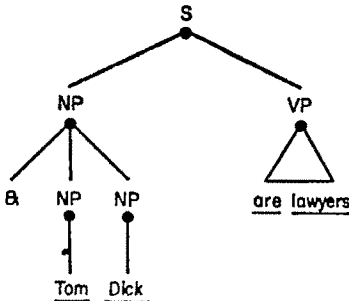
- (84) $NP \rightarrow NP^n$
 $VP \rightarrow VP^n$

Thus, for example, Dougherty's grammar assigns different deep structures to (85) and (86),

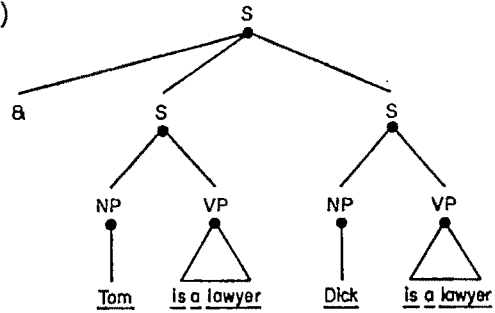
- (85) Tom and Dick are lawyers.
 (86) Tom is a lawyer and Dick is a lawyer.

namely (87) and (88) respectively.

(87)



(88)



Dougherty tries heroically to account for the syntactic and semantic properties of sentences containing NP's and VP's by means of *features* that qualify NP and VP nodes. These features are [$\pm$ individual], [$\pm$ total], [$\pm$ disjunctive], and [$\pm$ negative]. Though he cautions that these features are to be understood as operationally defined by the grammar, Dougherty ([5], p. 868) gives them the following intuitive content. The feature [$\pm$ individual] "emphasizes the independent, individual action of each of the constituents of the conjunction" and "suggests that a conjoined sentence paraphrase exists"; this feature is associated with such adverbs as *alone*, *singly*, and *individually*. The feature [$\pm$ totality] suggests that the dominated constituents behave "like the elements in a series bound together by plus signs"; it rules out occurrence with the [$\pm$ individual] adverbs mentioned above. The feature [$\pm$ totality] "emphasizes that the conjunction is to be considered as a unity; the conjuncts act mutually"; it is associated with such adverbs as *together*, *simultaneously*, *en masse*, *at once*, and *in chorus*. The feature [$\pm$ disjunctive] or [$\pm$ negative] is given, perhaps because Dougherty thinks that these features wear their intuitive content on their sleeves. The feature complex [$\pm$ total, $\pm$ ind, $\pm$ disj, $\pm$ neg] becomes the surface quantifiers *all* or *both*, and [$\pm$ total, $\pm$ ind, $\pm$ disj, $\pm$ neg] becomes the surface quantifier *all*, but not *both*. The feature complex [$\pm$ total, $\pm$ ind] is not realized by any surface quantifier but is associated with such words as *disperse*, *homogeneous*, *heterogeneous*, *scatter*, etc. The complexes [$\pm$ total, $\pm$ ind, $\pm$ disj, $\pm$ neg] and [$\pm$ total, $\pm$ ind, $\pm$ disj, $\pm$ neg] become the surface quantifiers *either* and *neither* respectively (see [5], pp. 868 ff).

¹⁸ This statement is not entirely accurate. A small but interesting class of *surface* NP's are derived by a conjunction substitution transformation from underlying sentential conjunctions. See Dougherty [5], pp. 865 ff.

Dougherty's grammar makes many true predictions, but it also makes many false ones. For example, it predicts that, contrary to fact, the following sentences are all ungrammatical:

- (89) The rod and the brick are both homogeneous.
- (90) All metal rods and bars are homogeneous.
- (91) Neither the rod nor the brick is homogeneous.
- (92) Neither Tom and Dick nor Sally and Sue arrived simultaneously.
- (93) Either the boys or the girls acted in concert.
- (94) The German army and the French army each outnumber the English army.
- (95) Neither the boys nor the girls dispersed.

Dougherty's grammar is presented in two lengthy installments the first of which sets out the basic theory. In a footnote to the first paper ([5], p. 861, fn. 14), Dougherty says he "will be mainly concerned with coordinate conjunctions of semantically singular noun phrases... not... because they have any great intrinsic interest, but because it is coordinate constructions of this type which are relevant in choosing between (the rival hypotheses)... on grounds of empirical adequacy." Notice that our counterexamples (92)–(95) involve NPⁿs all of whose conjoined constituents are semantically non-singular (in [92], "Tom and Dick" and "Sally and Sue" are the conjoined non-singular constituents); had Dougherty attended more to such NPⁿs, the empirical inadequacy of his own theory would have been manifest. A footnote to the second paper proclaims an apparently more liberal *restriction of attention*, viz. to NPⁿs that have at least one semantically singular constituent,¹⁹ for *ostensibly* the same reason, viz. these are supposedly the NPⁿs relevant to deciding among competing theories. Not until the last two pages of the second paper do we find a candid admission that the theory yields false predictions for sentences with NPⁿs all of whose conjoined constituents are semantically non-singular.²⁰ Such counterexamples are not manufactured oddities amenable to a minor adjustment of the theory; they are garden-variety

NPⁿ sentences that go to the heart of the grammar.

Recursiveness or *nesting*, the quintessential trait of language according to transformational grammarians, undermines Dougherty's grammar. Consider the ill-formed sentence (96)

- (96) *(Neither [(Tom and Dick) or (Sally or Sue)] nor [John and Bill]) (arrived together or walked home alone).

under the reading indicated by parentheses. The ungrammaticalness of (96) is *independent* of the topmost connective of its surface NPⁿ and of the embedded "or." Call them what you will and multiply them as you might, no features realized *inherently* by these connectives could predict the ill-formedness of (96). Sentence (96) is ungrammatical because the deeply nested NPⁿ "Sally or Sue" cannot serve as the subject of the embedded VP "arrived together"; "Sally or Sue" cannot so serve because the underlying dyadic predicate "arrive with" requires two arguments. What this shows, then, is that Dougherty's grammar is at best a grammar of *non-nested* NPⁿs. Sentences such as

- (97) Tom, Dick, and Harry are lawyers and friends.

administer the *coup de grâce* to Dougherty's grammar. The latter accords (97) a deep structure having "Tom, Dick, and Harry" as NPⁿ and "are lawyers and friends" as VPⁿ. But as our theory of mongrel conjunction makes evident, the surface constituent "Tom, Dick, and Harry" plays radically different roles in (98) and (99).

- (98) Tom, Dick, and Harry are lawyers.
- (99) Tom, Dick, and Harry are friends.

Dougherty would mark the NPⁿs of (98) and (99) as [+ind, +tot] and [–ind, +tot] respectively. Thus no complex feature can consistently be assigned to the NPⁿ of (97). Sentence (97) cannot be generated by phrase-structure rules as Dougherty's grammar purports to do. It must be derived from (98) and (99) either by a conjunction reduction transformation or by a conjunction substitution transformation akin to Dougherty's.²¹

¹⁹ Dougherty [6], p. 320, footnote 24. Actually Dougherty misstates the restriction he obviously intends, saying that he will be concerned with NPⁿs such that "at least one of the coordinated elements is a semantically non-singular NP". The same misstatement appears in the one other place where Dougherty mentions the restriction, *ibid.*, p. 334. This suggests to us that the restriction represents a desperate *ad hoc* attempt by Dougherty to save his grammar in the face of garden-variety but wholly unforeseen counterexamples.

²⁰ Dougherty [6], p. 337 f. Curiously, Dougherty does not hesitate to apply his grammar to sentences that violate the restriction when such application yields true predictions.

²¹ See Dougherty [5], pp. 865 ff., for a discussion of conjunction substitution transformations.

VI. GROUP QUANTIFICATION

Where there is conjunction, there must also be universal quantification. So intimately related are they that in a finite domain a universally quantified sentence " $(\forall x)\varphi x$ " will be true just in case the conjunction " $\varphi a_1 \& \dots \& \varphi a_n$ " is true, where ' a_1 ', ..., ' a_n ' is a list of individual constants that exhausts the domain. To each of the five species of conjunction that we have identified, paradigms of which are given by (100)–(104),

- (100) Tom, Dick, and Harry are lawyers.
- (101) Tom, Dick, and Harry are similar (to each other).
- (102) Tom, Dick, and Harry are similar (to one another).
- (103) Tom, Dick, and Harry carried the piano upstairs.
- (104) Tom, Dick, and Harry are shipmates.

we should expect a corresponding species of universal quantification, an expectation borne out by (105)–(109).

- (105) All the men are lawyers.
- (106) All the men are similar (to each other).
- (107) All the men are similar (to one another).
- (108) All the men carried the piano upstairs.
- (109) All the men are shipmates.

Sentences (105)–(107) instantiate the three varieties of (universal) *distributive* quantification that answer to the three varieties of *sentential* conjunction. Since the theory of these three species of distributive quantification amounts to a straightforward extension of our theory of sentential conjunction, we will rest content merely to depict the deep structures of (105)–(107), viz. (110)–(112) respectively:

- (110) $(\forall x)(Mx \supset Lx)$
- (111) $(\forall x)(\forall y)(Mx \& My \& x \neq y \supset Sxy)$
- (112) $(\forall x)[Mx \supset (\exists y)(My \& x \neq y \& Sxy)]$

Lakoff ([14], p. 558) has recently offered linguistic evidence for "a group reading of quantifiers of certain sorts, the logical form of which is unknown." The quantifiers that Lakoff calls "group quantifiers" are the four types exemplified in (106)–(109). The lumping together of these quantifiers as "group quantifiers" is the quantificational analogue of Lakoff & Peters' lumping together of (101)–(104) as "phrasal." We shall ourselves reserve the term *group (universal) quantifier* for quantifiers like the one in (108); since it exhibits both group and distributive aspects, we

shall call the quantifier of (109) a *mongrel (universal) quantifier*.

One striking difference between distributive and group quantification concerns scope. Lakoff ([14], pp. 555 ff.) remarks that, unlike distributive quantifiers, group quantifiers are *scopeless*. Though insightful, this observation is misleading in four respects. First, it applies only to that one species of universal quantifier that we have called *group quantifiers*. Second, it does not hold at all for *existential* group quantifiers. Third, it is true of bonafide group universal quantifiers under only the first of the two alternative analyses given below. Fourth and finally, even with respect to this first analysis of group universal quantifiers, the observation must be qualified. This first analysis was adumbrated in §1 where we indicated that the logical form of (108) could be taken as (113)

- (113) $(\sigma x)Mx$ carried the piano upstairs.

or more schematically, that the logical form of (114) is (115)

- (114) $\psi(\text{all the } \varphi s)$
- (115) $\psi(\sigma x)\varphi x$

whenever (114) receives a group reading. The phrase " $(\sigma x)\varphi x$ " is a singular term that purports to denote; as such it has no scope. So, if one takes the entire phrase "all the φs " to be the group quantifier in (114), then group quantifiers lack scope. But if one means by the group quantifier of (115) only the operator " (σx) ," then group quantifiers have scope just as distributive quantifiers do. For example, the formula " $Mx \vee Wx$ " is the scope of " (σx) " in (116).

- (116) $(\sigma x)(Mx \vee Wx)$ weighed down the life-raft.

which is a paraphrase of one reading of (117).

- (117) All the men and women weighed down the liferaft.

By appealing to the alleged scopelessness of "group" quantifiers, Lakoff purports to explain why (118) and (119)

- (118) John and all the linguists are similar.
- (119) *John and every linguist are similar.

come out grammatical and ungrammatical respectively. The alleged explanation goes like this (see [14], p. 557). One of Ross' constraints on movement transformations forbids moving an "element into or out of one conjunct of a coordinate structure". Unlike "all," "every" can function

only as a distributive quantifier. But distributive quantifiers originate *outside* the coordinate structure "John and every linguist." So, no transformation can move "every" into "John and every linguist," for such movement would violate Ross' aforementioned constraint. On the other hand, the scopeless group quantifier "all the linguists" in (118) originates *inside* the coordinate structure in which it occurs; its appearance there thus violates no movement constraints. Were the foregoing a good explanation, the transparently grammatical sentence (120)

(120) Each dog and its master are similar.

would be ungrammatical, since (120) has a deep structure along the lines of (121).

(121) $(\forall x)(\forall y)(Dx \& Myx \supset Sxy)$

Moreover, it is not obvious that (119) is ungrammatical. Suppose that John resembles every linguist in that John uses his hypotheses to determine the putatively neutral data that are supposed to confirm them; to report this fact one might well use (119) with the reading given by (122).

(122) $(\forall x)(Lx \supset S_jx \& S_xj)$

McCawley ([20], pp. 526 ff.) has proposed what might be dubbed a *proto-analysis* of group quantification. It amounts to taking group universal quantifiers as singular terms denoting sets. For example, "all the men" in (108) is taken to denote the set of (contextually restricted) men, while "all the linguists" in (118) denotes the set of linguists. Presumably, phrasal conjunctions would be handled analogously. For example, "Tom, Dick, and Harry" in (103) would be taken to denote the set {Tom, Dick, Harry}. To turn this proto-analysis into a full-blown theory, one would have to develop it on the pattern of our own theory. And of course this could be done, for any theory whatsoever can be modeled in set theory. Nevertheless there are good reasons, both of conscience and system, for preferring our theory over its set-theoretical counterpart. Let us, however, begin with a bad one. It might seem that McCawley's proto-theory effects a rupture between (103) and (123),

(123) Tom carried the piano upstairs.

for (123) applies to an individual the very predicate that (103) applies to a set. However, merely by choosing a set theory that identifies individuals

with their unit sets, i.e., a set theory in which "Tom = {Tom}" is true, one disarms this objection completely.²²

As a matter of conscience, many philosophers resort to sets only in desperation. Their maxim is to avoid commitment to sets whenever and so long as possible. We have shown that mereological predicate logic affords a general theory of English NPⁿ constructions that makes no commitment to sets. Accordingly, philosophical conservatism favors our theory over any set-committed surrogate. Judged systematically, our theory wins hands down over theories, like McCawley's, that incorporate set theory; unlike the former, the latter are Gödel incomplete.

We turn now to mongrel quantification. To facilitate presentation of our theory, we let (124) abbreviate (125),

(124) y is a sum of φ s

(125) $(\forall z)[Oyz \supset (\exists w)(\varphi w \& w < y \& Owz)]$

so that (126) comes to abbreviate (127).

(126) y is a sum of φ s none of which is x

(127) $(\forall z)[Oyz \supset (\exists w)(w \neq z \& \varphi w \& w < y \& Owz)]$

Where Ω represents *shipmate*-type predicates, we assign to (128) the deep structure (129).

(128) $\Omega(\text{all the } \varphi\text{s})$

(129) $(\forall x)(\forall y)[(\varphi x \& y \text{ is a sum of } \varphi\text{s none of which is } x) \supset \Omega xy]$

Consequently, our theory assigns to (109) the deep structure (130)

(130) $(\forall x)(\forall y)[(Mx \& y \text{ is a sum of } M\text{s none of which is } x) \supset Sxy]$

which abbreviates (131).

(131) $(\forall x)(\forall y)[Mx \& (z)(Oyz \supset (\exists w)(w \neq x \& Mw \& w < y \& Owz)) \supset Sxy]$

The group *existential* quantification analogue of (108) is (132).

(132) Some of the men carried the piano upstairs.

We let (133) be schematic for sentences like (132).

(133) $\psi(\text{some of the } \varphi\text{s})$

Our theory assigns to (133) the deep structure (134).

(134) $(\exists y)(y \text{ is a sum of } \varphi\text{s} \& \psi y)$

which abbreviates (135).

(135) $(\exists y)[(\forall z)(Oyz \supset (\exists w)(\varphi w \& w < y \& Owz)) \& \psi y]$

²² For an example of such a set theory, see Quine [22], pp. 135 ff.

If there is at least one φ , (133) will come out true whenever (114) comes out true. But if there are no φ s at all, (133) comes out false, as does (114) in view of our convention (see footnote 3) about non-denoting singular terms. So on our theory, (114) implies (133) so long as ψ is a simplex predicate, just as " $(\forall x)\psi x$ " implies " $(\exists x)\psi x$ " so long as the domain of quantification is nonempty. Note, however, that where " ψx " stands for " $\sim Fx$," a complex predicate, (114) does not imply (133).

The distributive quantifiers can be defined in terms of each other and negation. For example, one can define " $(\exists x)\psi x$ " as " $\sim (\forall x)\sim \psi x$." But if we press the analogy between existential and group quantifiers by trying to define " ψ (some of the φ s)" by " $\psi \sim (\sigma x) \sim \varphi x$," we get an ill-formed expression in which a tilde modifies a singular term. This is not surprising since group universal quantifiers are related to phrasal conjunctions, whereas group existential quantifiers exhibit both phrasal and sentential aspects.

Note that the group existential quantifier has scope. It would be better, then, to rewrite (133) as (136),

$$(136) \quad (\text{Some of the } \varphi\text{s})\psi$$

or even more perspicuously as (137).

$$(137) \quad (\Sigma u\varphi u)\psi u$$

This rewriting allows us to take the scope of " $(\Sigma u\varphi u)$ " to consist of itself together with the wff that it modifies. Formation-rule clauses for group existential quantifiers are given by (138).

$$(138) \quad \text{If } \varphi u \text{ is a wff and } u \text{ is an individual variable, then } "(\Sigma u\varphi u)" \text{ is a group existential quantifier. If } "(\Sigma u\varphi u)" \text{ is a group existential quantifier and } \psi u \text{ is a wff, then } "(\Sigma u\varphi u)\psi u" \text{ is a wff.}$$

The Tarskian semantics of group existentially quantified formulas is straightforward. Consonant with it, we can read " $(\Sigma x Gx)Fx$ " as "some of the G s are F ," or less ambiguously as "some sum of G s is F ." If mereological predicate logic is supplemented with the iota operator, neither group quantifier need be taken as primitive. But in the absence of the iota operator, the sigma operator of group universal quantification must be taken as primitive. The group existential quantifier, however, can be defined in mereological predicate logic unsupplemented by any special operators. When the group existential quantifier is taken as primitive, one can

assign to (133) the deep structure (137) which is equivalent to (135).

We turn now to the second analysis of group universal quantification. Among its many virtues, this analysis closes truth-value gaps without recourse to special conventions governing non-denoting singular terms, it relates the group universal and existential quantifiers closely and naturally, and it dispenses with iota and sum operators. Rewriting (139) as (140)

$$(139) \quad \psi(\text{all the } \varphi\text{s})$$

$$(140) \quad (\Pi u\varphi u)\psi u$$

and providing formation-rule clauses patterned after (138), we analyze (140) as (141).

$$(141) \quad (\forall u)[(u \text{ is a sum of } \varphi\text{s} \ \& \ (\forall w)(w \text{ is a sum of } \varphi\text{s} \supset w < u)) \supset \psi u]$$

On this analysis, the deep structure of (108) becomes (142)

$$(142) \quad (\Pi x Mx)Cx$$

which is equivalent to (143).

$$(143) \quad (\forall x)[(x \text{ is a sum of } Ms \ \& \ (\forall y)(y \text{ is a sum of } Ms \supset y < x)) \supset Cx]$$

Like its existential mate, the group universal quantifier has scope and can be defined in unsupplemented mereological predicate logic. Moreover, although " $(\Pi x Gx)Fx$ " and " $(\Sigma x Gx)Fx$ " are not equivalent to " $\sim (\Sigma x Gx) \sim Fx$ " and " $\sim (\Pi x Gx) \sim Fx$ " respectively, the latter two formulas do imply the respective members of the first pair. Consonant with a Tarskian semantics for group universal quantification, we can read " $(\Pi x Gx)Fx$ " as "all the G s are F ," or more suggestively as "any sum of all the G s is F ." Of course, there will be exactly one sum of all the G s if there are any G s at all; the suggested reading takes account of the possibility that there are none.

Nursery rhymes offer much food for philosophical thought. Recall, for example, "Humpty Dumpty." It inculcates in children the Sellarsian thesis ([24], p. 221), borrowed from Aristotle, that "persons are, in the toughest of senses, single logical subjects: that persons are in no sense systems of logical subjects." For if a person, Humpty Dumpty, is reduced to a sum individual or even an organized system of individuals, his unity is lost beyond hope of recovery; all the king's horses and all the king's men couldn't put Humpty Dumpty back together again. The last sentence of the rhyme is a group

universal quantification. What better indication of the banality of group quantifiers, and hence of their centrality to our everyday conceptual scheme, could one hope for?²³

University of Pittsburgh

Received March 21, 1975

REFERENCES

- [1] RENATE BARTSCH, THEO. VENNEMANN, *Semantic Structures* (Frankfurt, 1972).
- [2] TYLER BURGE, "Truth and Mass Terms," *The Journal of Philosophy*, vol. 69 (1972), pp. 263-282.
- [3] TYLER BURGE, "Truth and Singular Terms," *Nous*, vol. 8 (1974), pp. 309-325.
- [4] DONALD DAVIDSON and GILBERT HARMAN (eds.), *Semantics of Natural Language* (Dordrecht, 1972).
- [5] RAY DOUGHERTY, "A Grammar of Coordinate Conjoined Structures: I," *Language*, vol. 46 (1970), pp. 850-898.
- [6] RAY DOUGHERTY, "A Grammar of Coordinate Conjoined Structures, II" (*sic*), *Language*, vol. 47 (1971), pp. 298-339.
- [7] LILA GLEITMAN, "Coordinating Conjunctions in English," *Language*, vol. 41 (1965), pp. 260-293.
- [8] ALVIN GOLDMAN, *A Theory of Human Action* (Englewood Cliffs, N.J., 1970).
- [9] NELSON GOODMAN, *The Structure of Appearance* (Second Edition: New York, 1966).
- [10] M. S. GRAM and E. D. KLEMKE (eds.), *The Ontological Turn* (Iowa City, 1974).
- [11] VIRGINIA HELD, "Can a Random Collection of Individuals be Morally Responsible?," *The Journal of Philosophy*, vol. 67 (1970), pp. 471-481.
- [12] HERBERT HENDRY and GERALD MASSEY, "On the Concepts of Sheffer Functions," in [16], pp. 279-293.
- [13] WILLIAM KNEALE and MARTHA KNEALE, *The Development of Logic* (Oxford, 1962).
- [14] GEORGE LAKOFF, "Linguistics and Natural Logic," in [4], pp. 545-665.
- [15] GEORGE LAKOFF and STANLEY PETERS, "Phrasal Conjunction and Symmetric Predicates," in [23], pp. 113-142.
- [16] J. KAREL LAMBERT (ed.), *The Logical Way of Doing Things* (New Haven, 1969).
- [17] HENRY LEONARD and NELSON GOODMAN, "The Calculus of Individuals and its Uses," *Journal of Symbolic Logic*, vol. 5 (1940), pp. 45-56.
- [18] GERALD J. MASSEY, "Are there any Good Arguments that Bad Arguments are Bad?," *Philosophy in Context*, vol. 4 (1975) pp. 61-77, published by Cleveland State U.
- [19] GERALD J. MASSEY, *Understanding Symbolic Logic* (New York, 1970).
- [20] JAMES McCRAWLEY, "A Program for Logic," in [4], pp. 498-544.
- [21] TERENCE PARSONS, "An Analysis of Mass Terms and Amount Terms," *Foundations of Language*, vol. 6 (1970), pp. 362-388.
- [22] W. V. O. QUINE, *Mathematical Logic* (Cambridge, 1951).
- [23] DAVID REIBEL and SANFORD SCHANE (eds.), *Modern Studies in English*. (Englewood Cliffs, N.J., 1969).
- [24] WILFRID SELLARS, "Metaphysics and the Concept of a Person," in [16], pp. 219-252.
- [25] WILFRID SELLARS, "Seeing, Seeming, and Sensing," in [10], pp. 195-215.

²³ I owe much to discussions with my Pittsburgh colleagues N. Belnap, J. Camp, M. Bennett, N. Rescher, R. Thomason, W. Zilonis, and especially W. Sellars who got me interested in conjunctive individuals.

II. THE VALUE OF TIME

JOHN LESLIE

I. COMPETING MODELS OF TIME

HOW will Time appear when seen through an axiarchist's spectacles? Axiarchists are those picturing the world as somehow ruled by ethical requirements.¹ So far as logic and the meanings of words are concerned, Reality might have been an ethical disaster. But an axiarchist could still hold that a person named "God" created the world *because* its existence was ethically required. More ambitiously, he might maintain that the existence of such a creator is itself due to his ethical requiredness. Again, it might be that "God" names not a *person* but only the principle that some ethical requirements *are themselves creatively effective*. In modern times, P. Tillich appears to have taught this; it has been taught for centuries. . . . Now, axiarchistic explanations of the world inevitably influence theories about the world's nature.

Let me illustrate this with a story. A kindly demon, impressed by the pattern displayed by our world in space and time, is still disappointed to find that in time it is chopped into infinitely thin slices, each annihilated just as the next makes its appearance. So he creates innumerable new worlds, their different patterns corresponding exactly to the patterns existing in ours at successive instants in our pasts. He further ensures that these worlds will never alter. They will remain frozen copies of situations which, in our world, time's tooth at once attacked. And then he brings them together, placing them at successive points along a line running at right-angles to their three spatial dimensions. After appropriately modifying our perceptual apparatus he invites us to admire his handiwork. In the four-dimensional continuum we find reproduced every detail of our lives. Looking down the new dimension we see "successive states" (i.e., cross-sections) of entities which are, in all but the precise way in which these "states" succeed one another, mirrors of our past selves. We perceive

the words and even the thoughts of these entities, and they again reproduce the patterns of our past words and thoughts. So some concern the passage of time: the fascination of a pendulum's swing, the exhilaration of action, the learning of tensed language, the pleasures of the unexpected, the bitter fear of dying. "Why," we exclaim, "these people have been tricked; they have been led to believe that they exist in a changing world when theirs will in fact always be changeless. Without knowing it, they are secure from ultimate extinction. How sad that no such trickery can have been practised on us!"

Though told in a disgracefully question-begging way, this tale may give some insight into the axiarchist's probable preference for the theory which makes time into what J. M. E. McTaggart christened a B-series. Following others, let us call this *the B-theory*. The B-theory views pastness, presentness and futurity as relative, so that an event cannot ever *in itself* have, say, the property of being present. It isn't merely that it cannot retain such a property throughout history (for nobody in his senses would deny this). Rather, *at no point in history* is presentness possessed by an event *intrinsically* rather than relatively. No happening can ever—however fleetingly—in itself possess this property, when "in itself" indicates that something other than a simply relational property is in question. [Compare how no man in himself ever has on-top-of-ness or far-from-ness. He may be on top of the Great Pyramid and far from London, but he can still have underneath-ness (relative to a cloud) and nearness (in relation to the Nile).]

The B-theory does not dismiss time as an illusion. Some events really are here—i.e., really do have nearness relative to me as I write this. Likewise, some really possess *nowness*; but what the theory holds is that they possess it very much as they possess being-to-the-left-ness. The suggested interpretation of "Jam today!" is, roughly,

¹ I introduced the term *axiarchism* in "The Theory That the World Exists Because It Should," *American Philosophical Quarterly*, vol. 7 (1970), pp. 286–298. Ethical requirements are described further in "Ethically Required Existence," *American Philosophical Quarterly*, vol. 9 (1972), pp. 215–224.

"Jam in the temporal vicinity!". "Past," "future," "present," are treated by the B-theorist as inessential words, translatable in terms of earlier-than-ness, later-than-ness and simultaneity. Should the translation meet resistance as not capturing "the real meaning" of the words, the meaning which common sense has encapsulated in ordinary language, then he may abandon them.

The B-theory, in effect, makes position in time into a straightforward affair of what is indicated by clock faces and dates in history books, just as spatial position is by grid references and entries in surveyors' notebooks. The flow of events is not mysteriously very different from the flow of a curve across an artist's canvas. Men travel onwards through time much as a highway travels onwards across a landscape, and after death recede in it much as a target recedes from riflemen each stationed farther off than the last. The world's organization develops, but so does a design develop along the length of a carpet. Change exists, atoms complexly interweave, but so are there changes (i.e., differences) in the pattern displayed by the carpet's successive cross-sections, perhaps through the interweaving of variously coloured threads. The instantaneous present is not denied, but it is as much an abstraction as these infinitely thin cross-sections.

The alternative, the A-theory, views presentness or nowness as a property which each event comes to possess intrinsically at some point in time.

In so far as A-theorists allow past and future events to have real properties, they think of pastness and futurity as possessed similarly. But so very seriously do they take the metaphor that the past is now cremated, the future not yet born, that they have some problem in saying *by what* any properties of pastness or futurity could be possessed. Past and future events are in no readily understandable sense "there" to possess them, if the A-theory is right, for it conceives nowness as an event's winning an existence which earlier was and later will be *absolutely* lacking to it. Lacking to it, that is to say, in more than just its failing to exist inside earlier and later cross-sections of space-time, must as New York fails to exist inside Russia. An event which has become past has therefore been annihilated. Existence, says the A-theory, is possessed by it neither in the temporal "here" nor in any tem-

poral "elsewhere"; it is not *existent in the past* but merely *was, in the past, existent*. The temporal position which it once filled can still be followed in thought as it recedes forever further into the past, but there can be nothing *in* that position, just as there can be nothing in temporal positions which are future—though there *will be* something in them. [For otherwise the A-theorist, in order to cling to his central claim that nowness is absolute, would be forced to picture it in the preposterous fashion attacked by D. C. Williams in "The Myth of Passage":² namely, as some sort of glow racing down events which are constantly real. Or constantly existent; for what could "real" mean, when applied to events, except existent, i.e., taking place? Or is taking place, as Williams witheringly asks, "a formality to which an event incidentally submits"?]

What is clear is that A-theorists are united in viewing position in time as very different from position in space. Being-later-than, being-simultaneous with, being-earlier-than, are certainly all only relative; but they are supplemented by a none-relative futurity, presentness and pastness which each event comes to possess in this order. The pattern of events does not develop merely in historical time (calendar-time, clock-time), revealing its successive parts as a landscape reveals its houses to observers at successive mile-posts. For historical time itself develops in this sense: that nowness, alias Becoming or the gift of existence, is constantly transferred from the events of one second to those of the next. Every event thus has a place on two time scales: on a "B-scale" bearing such markings as "Two o'clock on June 18th, 1815" and also on an "A-scale" whose main divisions are "As yet absolutely non-existent (future)," "existent (present)," and "was existent but now absolutely not so (past)." Though on the first scale its location never alters, it moves along the second in a manner not in the least dependent on the positions of those who think about it. Events could in this manner vary their temporal characteristics even if thinkers never existed.

II. WHY AXIARCHISTS MAY PREFER THE B-THEORY

Now, the B-theory can seem far the better of the two for the purposes of any axiarchist—any

² *Principles of Empirical Realism*, chapter 15.

man who tries to account for the world's existence by reference to its ethical requiredness. Williams (*op. cit.*) accurately reports the dismay filling so many when they think of the time of the A-theory, "the flood on which the oldster wakes in the night to shudder at its swollen black torrent cascading him into the abyss." In contrast the B-theory, while not denying our limited life-span, at least throws doubt on the reasoning behind "that sad anxiety to *keep going* futurewards" which contrasts so strangely with our comparative indifference towards spatial restrictions. It could perhaps be thought a shame that we cannot live at every instant in historical time, and it is an evident evil that some have lives as brief as they do. But why look on these facts as disastrously different from the facts that we cannot live at and experience every point in space at once, and that the world of some of us is a hospital ward? If, as the B-theory contends, position in time is as relative as position in space, then temporal limits to experience are very comparable with spatial ones, and death is much like the boundary of your visual field. The dying B-theorist is not tortured by any wish to have his life all over again, precisely as before, because to him this would be almost like asking to be drunk enough to see double. On his view, his future absence will no more involve him in absolute inexistence than does his present absence from the room next door. Certainly he will not be *then*, but neither is he *there* next door or on Mars or Jupiter.

Note (1) In the year 3,000 it would be thoroughly misleading to say that a man whose life really did end in 2,000 "exists now." And the B-theory does not dispute this. It no more doubts that men of the sixteenth century are absent from the present day than that Venice is outside the Sahara. What it doubts is their absence from existence. (2) When the B-theorist calls it inappropriate to pity men of past centuries just for the fact that life is a good which they do not have *now*, this isn't to deny that they should be pitied because less able to mend broken ankles or less provided with transistor radios. Nor is it being said that pity is pointless because nothing can be done about their sad fate, let alone because (as Plotinus strangely recommends to our attention, should we have any fear of losing the gift of life) "they simply aren't there to have any sad fate." Suppose that a man dying after a long and happy career envies

a young girl with leukaemia on the grounds that she still has some months to live. The B-theorist who called this inappropriate wouldn't be saying merely (a) No use crying over spilt milk; (b) Be content that you had your turn and that it was a good one; (c) And bear in mind that being dead really isn't in the least unpleasant. What he would chiefly have in mind would be the comparison with the man who, living a happy and active life in France, envies any vaguely contented prisoner's on Devil's Island just for being somewhere where he himself is not. (3) True enough, temporal limitations on our experiences are in some ways more deplorable than spatial ones. It could be exciting to move at will from epoch to epoch as you can from Paris to Marseilles. But inability to do so could never inspire the kind of terror which limited life-span so often evokes—a terror which, whether or not unreasonable, is still very *understandable* in those holding the A-theory belief that temporal absentness is not a purely relative affair like being spatially absent, i.e., distant. (4) It would be odd for an *A-theorist* to worship as benevolent any god who intended simply to let life die out with the heat-death of the one and only cosmos that he would ever create. On the other hand, it would be utterly grotesque for a *B-theorist* to join J. B. S. Haldane in staving off the fear of the grave with the hope that, given sufficient time, our cosmos would revive itself from its heat-death and set itself going again just as before, so that he would live the same life once again.

All this amounts to nothing so monstrous as that doctors should accept a child's death complacently; for where have I called for complacency when partial blindness and paralysis narrow someone's experiences of space? Still, the B-theorist's proposal—that people of past centuries are no more *absolutely inexistent* than people of distant countries—may help to make an axiarchistic world-view more acceptable. None of the past *exists now*, not even in a ghostly fashion, where "now" means "here in historical time"; but neither, when I am in Canada, is Australia "existing here in space." And the greater part of death's terror surely stems from a belief that lives, when they are past are absent not just relatively but absolutely.

But doesn't an A-theory universe offer compensations? Surely A-theory changes—the strictly non-relative way in which successive situations

take on and cast off existence, the charms of novelty—are worth the pangs of recognition that there will come a day on which we shall be absent not just *from the set of things having nowness at that date* but *absolutely*? . . . No; and not because this errs about which of two goods is the greater, but because it makes no sense whatever. Three separate mistakes might be detected in it: (I) It seems to suppose that the B-theory contains no place for the delights of variety, the surprise and challenge of novelty, the plain fact that the world's pattern develops with us as we grow older. (II) It perhaps compounds this first error with another yet more glaring. "The B-theory world," it is reasoned, "contains *no change of any sort*" (error number one); "and therefore anybody in such a world, even if initially interested by it, *would soon change to being bored*." (III) It seems to forget that a thing is what it is, not what it was. For when he remembers this, mustn't every A-theorist accept that the value of a situation depends only on what that situation itself is and not on the truth that what preceded it—which, says he, is absolutely nonexistent—was different? Novelty as conceived by the A-theory, a bare way in which events are other than those which once existed but now have been annihilated absolutely, is a strange candidate for the role of something adding value to the world! The B-theorist certainly doesn't deny that Reality is different (and therefore novel) at each successive clock-second, and so has the richness of variety. He is all the more sure of this *because* he rejects the A-theory; for are you *rich* when a million gold coins pass through your hands, but each must leave them before they can grasp the next? Novelty, Variety, the throb of Becoming, are to the B-theory simply "the existence of an entity, or kind of entity, at one time [date] in the world continuum which does not exist at any previous time [date]. No other sort of novelty than this is discoverable—or desirable."³

It is the A-theorist who robs novelty of its worth, leaving axiarchism with no satisfactory answer to the question: "If the world somehow owes its existence to its ethical requiredness, then why are there changes in it? Why wasn't it so good from the very start that any alterations could only be deteriorations?" The B-theorist, in contrast, can *add* the values of situations in all

ages when estimating the strength of any ethical need for a cosmos like ours. He doesn't see any dimension of time as allowing only for *variations* in Reality's value. But remember that he needn't therefore be taken as claiming that all events exist *simultaneously* or *eternally*, because on the most natural reading this is a claim that all exist at, respectively, *one and the same date* and *every date*, which would be like claiming that all are at one and the same or at every spatial position. What is instead claimed is that at any date the absence of the events of other dates is simply their failure to occupy the temporal neighbourhood.

III. HOW THE B-THEORY SURVIVES THE TEST OF EXPERIENCE

But does the B-theory really square with experience? Could occurrences in the far future, say, be thrown onto the screen of the here and now by any physically possible means, much as far galaxies can be made available for our inspection by a telescope?—Well, I dare say they couldn't be; but this is too simple as an objection. Why should the B-theorist think that if events of the year 3,000 had absentees only relatively, not intrinsically, then light rays (for instance) would bridge the gap? You might equally well raise an egg; comment on its nature; drop it; comment on its new, smashed nature; and conclude that since a change has occurred the B-theory must be wrong. Success is unlikely when you point to details of the world which the B-theorist himself accepts, then exclaiming "I refute him thus!". The B-model of time is plainly intended as modelling *change*. To demonstrate that two cross-sections of an egg's temporal spread really can differ, or that events which will be seen in the future cannot be seen now, misses the point.

The B-theory's immunity to all such simple demonstrations can perhaps best be shown by a restatement of the tale about the demon. Suppose that the A-theory correctly represents time's character. Nowness, the gift of actual existence, really is *transferred* from each situation to the next. The past and the future have an absentness from existence which isn't merely absence from the clock-time "here." Despite these truths about what is for argument's sake our actual A-world,

³ Williams, *op. cit.*; my insertions in parentheses.

isn't it *logically possible* for there to be *another* world, a B-world, in which the patterns of successive situations—their structures, the scientifically describable relationships between their parts—mirror the patterns which succeed one another in ours, though in this other world they succeed one another in B-theory time only? Might it not be that the patterns of this other world came into existence all at a single instant in the kind of time, A-time, in which our world is changing? Mightn't they even exist for ever in A-time? And as the patterns of this B-world match those of ours, what pattern of laboratory data could convince us that ours really is an A-world? For couldn't the other world have a space-like fourth dimension—let us call it a B-time dimension—along which the different patterns succeeded one another? Couldn't an egg (or, if you insist, a four-dimensional quasi-egg) approach the surface of a planet (or quasi-planet) along this dimension, much as two printed lines approach each other at points successively closer to their intersection? And couldn't laboratory apparatus record (or quasi-laboratory quasi-apparatus quasi-record) this development (or quasi-development—since *for argument's sake* our world of "genuine developments" is an A-world?

To imagine that the patterns of the other world all came to exist just when the last second of 1974 become *now* "in the A-time in which our own world is changing" is not to suppose, absurdly, that all the events of that other world came to exist together in the three-dimensional cross section of *that other world* which people *there* would consider as "the situation at the start of the last second of 1974." It supposes only that an observer in our world, if suitably mentally and physically equipped, could have detected the sudden coming to exist of a four-dimensional manifold—a manifold whose existence seems to me a definite logical possibility.

Yet don't our own experiences reveal what no laboratory records could? Williams (*op. cit.*) argues that inability to experience more than one Now cannot prove that other Nows have an absolute absentness, for it can be held that "there is no more a single rolling Now than a single rolling Here along a standing line of soldiers, though each of them has a vivid presentment of his own here." But doesn't this forget that we might have a vivid presentment of the A-nature of time as it rolled through each hun-

dredth of a second? Mayn't we be directly conscious of the very immediate future as it approaches or is created, or at least of the very immediate past as it recedes or undergoes annihilation?—No; not if the A-theory itself is to be believed. For this theory says that past and future events are absolutely inexistent. Any supposed consciousness of their approach or recession can therefore at most involve vivid models of them, models brought into the Now by memory or by anticipation. Our awareness of change needn't always involve such models; (nerve cells in the eye, for instance, react with special vigour to movements, and some of our consciousness of change may stem directly from this fact); but even when it does involve them there is nothing in this which couldn't be included in B-theory descriptions of the world. If, say, a memory-model of a past event is compared with a thumbprint on wax, then the B-theory can deal with it just as it can with any changes which wax undergoes. And such comments apply even to the very immediate past and very immediate future. If something allegedly in the immediate past were in fact—as suggested by some in their talk of Duration—only fading away, then it wouldn't be in the past at all; it would be a faint element in the present. And no experience of a fading could prove that whatever was fading would in the end take on inexistence absolutely, as distinct from becoming inexistent inside various clock-time cross-sections of a B-theory world.

To repeat: All this does not suggest that our consciousness of change, of the world's dynamism, is illusory. The point is simply that the data of consciousness cannot favour the A-model of change above the B-model. All the signs of time's reality which the A-theory can describe—the heard thunder of a passing train, the felt gust of its passage, the dizzying sights and movements of a switchback ride—can be incorporated into a B-theorist's story. This is true even of our feeling of being confined to the present minute and able to peer beyond it only through the dusty lenses of memory and prediction. The B-theorist in the dentist's chair can well say, "Thank goodness the pain is over *now*." Not just because he is glad that the pain's temporal spread is limited, like the spatial stretch of a tea-stain, but because he may not even know precisely which tooth had ached, let alone feel a compulsion to cry out in agony. The *practical* question, *What patterns do our experiences bear at par-*

ticular clock-times? must not be confused with the philosophical questions, *What exactly is the status of those patterns? Is the gift of existence transferred from one to another, as the A-theory says, or are presentness, pastness and futurity only relative?*

Naturally, what goes for our patterns of experience goes also for our patterns of argument. In a B-world you would still find A-theorists displaying their assumptions during one clock-minute and drawing their conclusions during another. Again, people in the two kinds of world would find it exactly equally hard to refer to things unambiguously without relying on tensed language. Any patterns logically capable of existing in absolute isolation are also capable of existing without such isolation. The patterns presented by successive cross-sections of a block of wood can remain the same whether that block is in its original state, or sliced up, or sliced up and then glued together again.

We must however handle this analogy cautiously, because B-theorists often dislike being painted as defenders of a "block" in which events all "exist together." A. Grünbaum⁴ condemns "the pseudo-image of the 'block universe'" with its implication "that past and future events co-exist with those that are present"; though he champions "the objective *becominglessness* of physical events", the belief "that *nowness* and thereby *pastness* and *futurity* are mind-dependent" (i.e., relative to the points in time from which events are viewed). Likewise Williams, complaining of "utter misrepresentation" when he is accused of presenting us with "a block universe," nonetheless maintains that "the total of world history is a spatio-temporal volume," a "four-dimensional fabric of juxtaposed actualities." For, he explains, the statement that an event "not present in time [i.e., not in the temporal present, the temporal 'here'] nonetheless exists is no more contradictory than that one not present in space [i.e., not *here* in space] nonetheless exists."⁵... But how can a B-theorist hold that talk of a block universe misrepresents him, yet that the universe forms a *four-dimensional volume* in which events failing to exist at the present instant *nonetheless do exist*? How can another maintain that *becoming* (*coming to exist*) *must always be relative*, yet recoil from the

statement that events throughout history co-exist? It seems that these writers run one risk of misleading us in order to remove another. What they deny is that all events could exist, like the cross-sections of a block of wood, "at the same time" or "for ever," when this means at a single clock-second or at all clock-seconds. A *block universe of co-existing events* must be disowned if by this is intended a set of events which exhausts all history yet extends in space only, not in time. But if what is intended is a becomingless four-dimensional fabric of juxtaposed actualities, then that is what the B-theory is all about.

Similarly with whether the B-theory "spatializes time." Accused of supposing that we see across time as across space, or that we move to and fro through time just as we please, or that we can find no difference between traveling and growing old, B-theorists have reason for annoyance. Yet they do make time space-like with respect to the relativity of "what is present": the temporal "here" behaves much like the spatial "here." Again they appear to make it at least logically possible that, as Williams expresses it, "a single human life should lie thwartwise of the manifold, with its conscious stream perhaps running along somebody's garden path. . . . True time travel, then, is conceivable after all. It would consist of a man's life-pattern running on an abnormal heading across the manifold." So even if time is not to be viewed as "a dimension of space," there may be no single time dimension clearly separable from three spatial ones in the four-dimensional continuum. . . . The fact is that "space-like" is discouragingly vague. The B-theorist's safest comment is that even if it may be somehow objectionable to speak of successive happenings as positioned along a space-like fourth dimension, the most that is lacking to such positioning is space-like relations between these happenings; for it is not as if *before* introducing such relations a playful demon would have first to create future happenings and re-create past ones, as the A-theorist would think necessary. However, it seems increasingly plausible for B-theorists to treat *just how the direction of time runs* as an issue on a par with that of *in exactly what direction the grain of the wood runs* along a plank. On such an issue there can be no uniquely correct answer.

⁴ In section 4 of "The Status of Temporal Becoming" from *The Philosophy of Time*, ed. by R. M. Gale.

⁵ *Op. cit.*; my insertions in parentheses.

IV. CONTINUED EXISTENCE AND CAUSATION IN A B-THEORY WORLD

"But," it might be objected, "doesn't all this overlook something crucial? Perhaps there is logical room for doubt over whether it was *our actual selves* who in the past experienced the events which we remember, or merely *other selves* joined to us much as one yard of railway line is to the next, selves who pass their brain-patterns along to us in much the sense in which any cross-section of a tree-trunk passes its pattern along to neighbouring cross-sections. Yet the latter alternative is inherently very implausible. Stop to consider how a tree-trunk or a railway line gets its structure. Don't *men who persist in time* plan and construct neighbouring sections of railway track, ensuring that each flows smoothly into the next? Don't atoms persist in time, so becoming subject to forces whose prolonged action organizes them into trees? Yet in what super-time would any organizing factors persist and work, manufacturing the four-dimensional threads which, says the B-theory, have three-dimensional atoms as their cross-sections? Or how, in the absence of organizing factors, could the successive cross-sections of any thread flow into one another so smoothly, and why should the four-dimensional filaments forming each thread—filaments whose three-dimensional cross-sections we should recognize as electrons, protons and other such particles—interweave in a way so reminiscent of the filaments in man-made threads? A pinch of common sense shows that history is acted out by entities enduring in a time which really flows, the time of the A-theory. The world's substance isn't duplicated endlessly along any 'time dimension' rather as the successive stones of the Great Wall duplicate one another with each yard of its 'progress' across China. Rather, the present perpetually preys on the past for the stuff of existence—stuff which, once arranged otherwise, is now not *supplanted by other stuff* but merely released from its previous patterns and rearranged. Hence for yesterday, today and tomorrow to lie juxtaposed on a fourth dimension would be like having three one-ton statues made with a single ton of bronze!"

Reasoning on much these lines is probably behind the man-in-the-street's acceptance of the

A-theory. [That he does accept it in some vague form can be deduced from his quite usual pity of the dead however long and happy "the lives which they once had," or occasional envy of them, however long and miserable those lives, because "for them it's all over."] (I) It certainly begs the question of how to define personal identity when it alleges that the B-theory makes "other selves" experience the past states which we like to think of as ours. (II) "Inherent implausibility"? Physicists today tend to be content with mapping reality as a collection of thread-like "world lines" whose serpentine extension through space-time is viewed as no more "inherently improbable" than any failure so to extend. Why should a lengthy rod, so to speak, be any more *a priori* unlikely than a disk? (III) To the axiarchist, however, the reasoning's main fault will be its blindness to his theory. Suppose it is declared that our cosmos, if it is to "run itself"—i.e., develop in a law-regulated way without external aid—must positively *run* as the A-theory maintains; must be "dynamic," not "static" as on the B-theory view; must contain things which pull or hit one another around in full-blooded A-fashion instead of existing as cross-sections of intertwined four-dimensional threads. As a start: Doesn't this question-beggingly deny that the B-theory has a place for whooshy thumpy dizzy-making dynamism? But equally important: Does the dynamic sound of a hammered nail, the energetic feel of muscular effort, truly provide us with deep insight into causation? Research papers in physics rarely use the word "cause," preferring to discuss the differential equations which our cosmos follows—equations which could as well be followed by a realm of four-dimensional threads as by any "strictly three-dimensional" scheme of things which, although containing "always the same stuff," nonetheless underwent the transformations perpetually imposed on it by a non-relative Becoming. But if (as I have argued elsewhere⁶) we do need an explanation of why the differential equations are followed, then the axiarchist can supply it. He can argue that events obey comparatively simple laws, laws of the sort which scientists can discover, because a world obedient to such laws has greater goodness—greater ethical requiredness—than any phantasmagoric system whose events are unpredictable. This ethical required-

⁶ "Does Causal Regularity Defy Chance?", *Idealistic Studies*, vol. 3 (1973), pp. 277-284.

ness influences the actions of a divine creator, as was argued in Berkeley's *Siris*, or else it is perhaps itself creatively powerful.

Many physicists much prefer the B-theory. First, they find it significant that we cannot always decide firmly whether it was space alone, or space and also time, which divided two events. Is it likely that physical laws would hide a rigid distinction between time and space if there really were one? Hermann Minkowski, impressed by how a four-dimensional map of history could please all observers only if dimensions of space and of time were not marked on it, concluded that "space and time separately are doomed to fade away into the merest shadows and only a combination of the two will preserve any reality."⁷ This found favour with A. Einstein, who judged that "it appears more natural to think of a four-dimensional existence, instead of, as hitherto, the evolution of a three-dimensional existence."⁸

Second, recent talk of "particles going backwards in time" supplies a perhaps even more impressive reason for favouring the B-theory—a reason more clearly supporting J. H. Jeans when he compared space and time to the length and breadth dimensions of a cricket field in which the ball "knows nothing of these distinctions: it goes where it is hit, directed only by laws which treat the field as an indivisible whole."⁹ In many situations physicists have to suppose either a process in which a particle somehow helps to conjure into existence two others, with one of which it then collides, or else, with J. A. Wheeler, R. P. Feynman and numerous followers, that *a single* particle has made a spatio-temporal zig-zag of a type thoroughly familiar from other cases in which a sudden acceleration is followed by sudden deceleration—though in this instance the zig of acceleration is great enough to give it a temporal path which A-theorists, for whom the past is in no sense "there" to be in any sense traveled to, will find impossible to describe. (1) It is, of course, most usual to define "travel" in such a way that we never really travel even *forwards* through time. But a fuss over this world would miss the point, which is that A-theorists cannot see a

single particle here, not even if you talk of zigging rather of traveling. Again, there is no need to make a fuss for or against the proposal that "from the particle's viewpoint" there is no question here of anything other than a steady progress forwards through time. (2) A protest often heard is that, were travel backwards through time a possibility, then we should face some variant on the logical paradox confronting time-travelers in science fiction: that they might seem able to prevent their own births. However, physicists very much doubt that the natures so far envisaged for backwards-traveling particles would permit the setting up of situations of this paradoxical type; and even if they did permit it, this might show only that they had been somewhat wrongly envisaged. The Feynman interpretation of particle movements does not suggest that backwards-traveling particles perform the nonsensical feat of "making the past other than it actually was when they started their backwards travel." They simply make it other than what it would have been if backwards-traveling particles hadn't formed part of it. It would certainly be logically absurd for an event to yield a causation-circle, involving past happenings, whose outcome was that these happenings happened otherwise than they actually did, so that the event itself didn't occur! A natural law might perhaps set up many other causation-circles in which each event's form was influenced by that of the rest, but it clearly could not set up causation-circles of this logically preposterous type. However, it is only inside the area of logical possibilities that a natural law can determine one sequence of events rather than another. Outside that area, the law would not so much 'break down' as become inapplicable. So the most a physicist would be forced to admit would be that certain laws permitting backwards time-travel would perhaps occasionally seem applicable to situations in which this applicability was in fact only apparent, other laws then applying in their stead. But these other laws could always be treated as codicils attached to the first laws—which is equivalent to dropping all talk of occasional inapplicability.

⁷ Address to Eightieth Assembly of German Natural Scientists and Physicians (Cologne, 1908).

⁸ *Relativity, the Special and the General Theory* (New York, 1920), Appendix 5.

⁹ *The Mysterious Universe* (New York and Cambridge, 1932), chapter 4.

V. ACTION, CHOICE, EXPERIENCE OF TIME'S FLOW

(I) The B-theory in no degree destroys the point of decision-making and effort. Since a B-world can include the same events as any A-world, later situations (wars, for example) can reflect the natures of earlier ones (say, diplomatic blunders) in the same way in each case. Still, the theory may well give problems to those who believe in libertarian free will. On their view, a man's decisions are not caused in all their details by the mental states preceding them. Moreover, the frying-pan of "total enslavement by causation" must not, they say, be replaced by the fire of "partial enslavement by chance." It must not be supposed that a man's freedom lies in any fact that the atomic particles making up his brain are governed only by statistical laws. Rather, men have a freedom which somehow escapes both these alternatives, a freedom such as no computer could ever have. Despite the scores of books and articles devoted to proving the reverse, such freedom is sometimes thought essential to morality. Now, it might well be held that the successive patterns of any world of libertarian free will would at least have to *come into* existence in the one-by-one fashion described by the A-theory, even if these patterns, when once in existence, accumulated to form an ever-lengthening four-dimensional sausage instead of each losing its existence as the next appeared. For a libertarian might reason that there would be no possibility of saying, prior to some free decision, that this decision or its effects had absentness only of the relative kind which the B-theory describes. You cannot "really decide" whether to smoke tomorrow, in the libertarian's preferred sense of "really decide," if it is in any sense already certain that the smoke curls upwards just a little further along a time dimension. If this reasoning is as forceful as I think, some axiarchists will feel compelled to sacrifice the B-theory out of respect for libertarian freedom. Perhaps they will introduce a life-after-earthly-death to offset this major sacrifice. My own preference is for denying that libertarian freedom would be particularly attractive.

(II) Again, there is no special need to ask how, in a B-world, we could get our feeling of time's "directedness." The B-theorist has no wish to dispute that some events really are earlier than others, or that ordinary language allows us to speak of a change *from* earlier events *to* later

ones, because no conceivable evidence—of laboratory experiment, of private awareness, of language or of how language is learned—could force acceptance of the A-theory. Absolutely any patterns which an A-world contained could be included in a B-world also. *Whatever* our reasons for talking about time's arrow, they would be present in both worlds equally.

It is all the same tempting to consider what these reasons might be. A standard answer is that earlier-to-later is the direction of entropy increase, often described as a gain in the "probability" or the "disorderliness" of successive distributions of the world's energy, which gives rise to the practical irreversibility of many processes. Most if not all are in theory reversible; but we can be sure that the majority of large-scale processes will not in practice reverse themselves. A wineglass dropped from a height is vastly unlikely to reform itself from its far-scattered atoms, even if in infinite time their heat-movements could produce this result. Now, the transactions which occur in our brains and our perceptual systems are among countless processes relying on entropy's increase. The energy redistribution essential to these transactions—processes of controllable information-transfer which allow us to think and to remember events, so that we become conscious of change—must be redistribution such as occurs in the time-direction in which the increase occurs. You cannot remember anything unless information about it can be transferred in a controlled way through any mechanisms involved in the remembering; hence memory is always of events in the time-direction in which entropy is lower. (1) Of course we might make it a matter of definition that *memory* can only be of events in this direction, if one grants that it is the direction in which *the past* lies; but this verbal manoeuvre is of little interest. For it isn't *memory so defined* but *controlled transfer of information* which is really in question. (2) But how can increase in the world's disorder yield the orderly workings of perception, thought and memory? The answer is that living mechanisms, like man-made ones, use large increases in disorder to produce limited increases in order. Power stations are run by the heat-energy scattering from their boilers to the rivers which cool their condensers.

Naturally the B-theorist, since he accepts into his world all the patterns which the A-theorist includes in his, need feel no less able than him

to explain entropy variations. Both equally may face a problem of why our cosmos at one point in clock-time possessed such low entropy. If this is a genuine problem, then perhaps the B-theorist is better equipped to answer it; for low entropy is often held to be evidence of Design, evidence supporting an axiarchistic interpretation of the world—and the axiarchist, I am arguing, has good reasons to be a B-theorist. But given that the universe did at one point possess low entropy, the laws governing entropy increase at other points follow with overwhelming probability from the basic laws of physics. So there is little force in A. S. Eddington's challenge to the B-theorist to "show your good faith by reversing the dynamic quality of time (which you may freely do if it has no importance in Nature), and, just for a change, give us a picture of the universe showing how, from man Nature in the course of ages achieved the sublimely simple structure of the amoeba."¹⁰ This challenge just demonstrates what any B-theorist admits: that the process of life's evolution relies on there being a direction in which entropy increases.

Remember: That there should be such a direction is overwhelmingly probable—on both the A- and the B-theory—by the fact that there is a point in time at which entropy is very low; for the reign of physical law makes it vastly unlikely to be equally low at all points. Now no doubt it is even more overwhelmingly unlikely that a world developing in A-time would decrease in its entropy like the world which Eddington challenges us to imagine. But he must not thrust his own A-theory slicer on us and then complain that space-time looks very odd when its slices are presented in (so to speak) south-to-north instead of north-to-south succession; for this is as unreasonable as slicing and presenting it east-to-west (i.e., parallel to any time dimension). When the B-theorist says that the slicer is not needed, this doesn't mean that you can wield it just as you will. Note: If looking at a B-world in the way Eddington so unfairly recommends, we should certainly see life's anti-evolution, people having memory-like knowledge of events which were going to happen, cars accelerated by the heat energy of their brakes, etcetera. But these couldn't be called *controlled processes*—processes such as we might expect to find at the basis of

thought and memory and experience of change. Instead they would be, as Eddington insists, fantastic matters of antichance, matters which could not have happened had the world's particles been arranged very slightly differently at the point in time at which we started our observations.

Whether or not the world's initial low entropy is good evidence of Design to support axiarchism, two attacks on this evidence fail. The first argues (with Haldane) that a universe, if existing for long enough, would sooner or later work itself far down the entropy scale by pure chance, just as air molecules in a cylinder would sooner or later group themselves all at one end. The second involves the rather similar reasoning that a universe sufficiently huge would be bound to contain large areas of low entropy, just as a sufficiently mighty army of monkeys strumming on typewriters would be producing a few Elizabethan dramas. But the difficulty with these lines of thought is that it is *immensely more likely* that the thinkers, their immediate surroundings, and an impressive stock of pseudo-memory-traces in their brains, should form chance oases in a spatio-temporal desert of disorder, than that an entire world low in entropy should form such an oasis.

VI. IS THE A-THEORY STILL VIABLE?

But is the A-theory really meaningful? Grünbaum complains¹¹ that the statement "It is 3 p.m. Eastern Standard Time now," if it doesn't say (as on the B-theory interpretation) that events of this clock-time have nowness relative to the saying or thinking of the statement itself, can express only "the utter triviality that the events of 3 p.m. EST occur at 3 p.m. EST." He detects no third option in insistence that at 3 p.m. nowness inheres in the events absolutely, since he is "totally at a loss to see that anything non-trivial can possibly be asserted by the claim that at 3 p.m. nowness (presentness) inheres in the events at 3 p.m.". A truly intrinsic property of nowness would be "unanalysable" and so presumably meaningless.—Yet is that true? Cannot we push an analysis at least this far: that A-theory nowness is a matter of a situation's *having an existence which could in the past have been said, and could in the future be said, to be absolutely lacking to it*

¹⁰ *The Nature of the Physical World* (New York and Cambridge, 1928), chapter 5.

¹¹ *Op. cit.*, section 3.

(rather than lacking to it merely inside particular cross-sections of space-time)?—Well perhaps this doesn't remove the difficulty, since it plunges us into McTaggart's Paradox concerning how the successive gain and loss of existence could be understood.

Time's reality requires every event to be first *future*, second *present*, and third *past*; but, said J. M. E. McTaggart,¹² no event can have all these characteristics together, so here we have a contradiction. And we cannot escape it by saying that each event has the characteristics at successive times; for time's reality has just been shown to involve a contradiction, and you cannot evade a contradiction by appealing to that contradiction itself.

The B-theorist can get around this, since an event could be future and also present and also past in the sense in which a village can be both to the left and to the right—because on which hand the village lies is relative to its viewers. But how about the A-theorist? He conceives time as involving *variations in the constitution of the totality of existence*, variations which are absolute rather than being matters of what can be judged "now" from varying vantage-points inside that totality itself. For the A-theory, therefore, a developed version of the paradox is:

- (1) The totality of existence is a totality having constitution ABC
and also (2) The totality of existence is a totality not having constitution ABC, since it has constitution XYZ instead.

Because time passes, A-theorists say, when (1) is the case and also (2) is the case thereafter. But, McTaggart would object, (1) and (2) contradict each other; and the word *thereafter* cannot improve matters. "Thereafter" is a time-word. A-theorists wish it to be given an A-theory interpretation. Hence it leads them into an infinite regress of attempts to escape the contradiction.

Yet isn't it enough for them to say that "the totality of existence" denotes in (1) something rather different from what it denotes in (2)?—Perhaps it is not enough; for we aren't here dealing with something like "the limited view from my window," which can of course denote different things for different speakers. It is *the totality of existence* which is in question. So to demonstrate that we escape saying that the

totality of existence *isn't* the totality of existence, we shall have to lean heavily on that word "thereafter." And this leads back into the infinite regress.

C. D. Broad rejected McTaggart's reasoning quickly on the grounds that, since there is in fact no contradiction in calling an event first future *and then* present *and then* past, there is no need to start on any regress to avoid a contradiction. I have however tried to go beyond McTaggart's actual words in a way showing that Broad's move, although available to any B-theorist, is by no means clearly available to his opponent. One cross-section of existence can be different from another; yet how could there be different realities each named *the totality of existence*? ... Admittedly, it might be enough to react to this very much as Broad did to McTaggart's seemingly more limited argument about individual events. Couldn't the A-theorist say that the *present-day* totality of existence can be (unlike the totality of what I can see from my window) an absolutely all-inclusive totality of existence, even though it doesn't include what *existed* yesterday and what *will exist* tomorrow? For his theory is that what existed and what will exist aren't in any sense "there" to be included in it. They do not lie just over something comparable to a spatial horizon. But the trouble is that McTaggart's general line of reasoning and this counter to it can both seem question-begging. The A-theorist: "We have ABC, and then we have not-ABC," McTaggart: "But what is meant by *then*?" A-theorist: "Well, ABC and not-ABC are descriptions both applying to the totality of existence, though they never apply *together*." McTaggart: "Make up your mind! How could they both apply if the first cannot apply if the second does?" A-theorist: "But that's what I've already explained! We have ABC and *then* we have not-ABC." McTaggart: "But what is meant by *then*?" And so on. I do not see that this succeeds in establishing firmly that the A-theory is wrong, though it surely points to a difficulty in it.

The A-theory also survives an attack mounted by Williams among others. "As soon as we say that time or the present or we move in the odd extra way which the doctrine of passage requires, we have no recourse but to suppose that this movement in turn takes time of a special sort.

¹² *The Nature of Existence* (Cambridge, 1921-1927), volume 2, book 5, chapter 33.

The history of the moving present then composes a new and higher time dimension, which cries out to be vitalized by a new level of passage, and so on forever" (*op. cit.*)—A-theorists could answer as follows. (I) the metaphor of movement shouldn't be taken too seriously, since what is intended is only a change in the status (will-be-existent, is-existent, was-existent) of various events. (II) The A-theory is in no way weakened by its recognition that time's passage *takes time*; for example, that it takes an hour for an hour to pass, just as you have to walk a mile in order to have walked a mile. The time in question is allegedly A-theory time, not something oddly extra. The accusation that the absolute movement of the present (i.e., says the A-theory, variation in the constitution of the totality of existence) would have to take time of a sort over and above any taken by the development of history (i.e., says the A-theory, variation in the constitution of the totality of existence) is patently unreasonable. [The A-theorist does not see the A-series (past-present-future) as entirely separate from and additional to the B-series (earlier-later). An event, in his view, is later than another only because it did absolutely gain and lose existence, does absolutely gain and lose it, or will absolutely gain and lose it, at a time lying in the future of that other event. Thus when Williams insists that "world history consists of actual concrete happenings in a temporal sequence; it is not necessary or possible that happening should happen to them all over again," A-theorists can comment, "We agree; and what is more the temporal sequence is the one the A-theory describes, a sequence of which the B-series running from earlier happenings to later ones is a mere logical consequence."]

VII. ACCEPTANCE OF TWO KINDS OF TIME?

Upshot: I concede the A-theory's defensibility. Its defenders might indeed avoid the obvious silliness of making events happen twice over, as well as the less obvious absurdity which McTaggart finds in it.

McTaggart might still complain that introduction of the little word *then* perhaps seems to remove the contradiction in "ABC and not-ABC" only because this word has not been given any clear meaning. However the A-theorist might answer that its meaning is sufficiently specified when we explain that it is the word which gets

around McTaggart's difficulty! Passage of A-time is to be defined as what removes the alleged contradiction between "The totality of existence has constitution ABC now" and "The totality of existence doesn't have constitution ABC now." A-time is that which permits real differences in how the totality of existence is constituted—differences which are not mere artefacts of each observer's position on a dimension of that totality itself. Strung together, the words in any such definition perhaps don't add up to anything meaningful. But they may do.

An equivalent description of A-time would be that it is the *logical possibility of absolute alteration*: alteration altogether different from any alteration in Europe's flatness as reported from Holland and from Switzerland. Once again, I am not at all sure whether "absolute alteration" is any more meaningful than "round square." But one thing seems evident. Were time to be understood in this style, then its flow wouldn't have to be accompanied by any *actual* alterations. Times would be two different times provided there were a *logical possibility* of the first one's being filled differently from the second. Time's passage could be modeled like this:

- (1) The totality of existence is such and such
- and then (2) There is a logical possibility that the totality of existence has become different from that described in (1).

With logical necessity, a thing is-what-it-is, but there may still be room for the A-theorist to declare, "There never is a logical necessity for anything to be now what it once was." Or, when the thing in question is the totality of existence, must this be like declaring that something is other-than-it-is?—I cannot say; my suspicion is that whether flesh could be added to the A-theory's skeleton of meaning is a matter not logically demonstrable. But if it could be added, then we should have to abandon the slogan that time measures change and is impossible without it. The A-theory suggestion, you see, is that there is no logical need for the totality of existence to remain *the same* as what it was, since it may change in A-time. Very well; but there would equally be no logical need for it to become *different from* what it was, even were A-time flowing.

Even in A-time, therefore, our world might be

the block or manifold which the B-theory portrays, since there is nothing in the A-theory to say that four-dimensional manifolds are logical impossibilities. It could make sense to ask how long in A-time such a manifold had existed and will exist, and to answer (for example) "for ever, and with all its events existing simultaneously." In effect, we might well acknowledge two types of time. Even were the A-theorist right about the logical possibility of absolute alteration, the B-theorist could also be right: right about the actual absence of such alteration. The only evolution which in fact occurred might be that which the B-theory describes, the difference between various cross-sections of a four-dimensional world. B-time, the time of clocks and of appointments, in which events of course do not exist eternally and all simultaneously, is in any case the only time of obvious practical relevance. And if the universe changed in B-time only, then A-time would be a mere philosopher's curiosity.

A thoroughgoing axiarchist will at any rate never believe in any period of A-time at which nothing existed, because the ethical need for good worlds would be the same at every point in such time. And he would have strong reasons for thinking that things would exist in A-time as parts of an unvarying four-dimensional whole or of several such wholes. Which is *not* a denial of variations in B-time.

None of this is clearly foreign to ordinary ways of thought. (I) Most people would agree that an unvarying system of things could undergo passage of time if at least one varying thing existed beside it. So why—they could well ask—would time suddenly cease to pass if that single varying thing were annihilated? Wouldn't it still be true that it might have continued to be there and to vary—i.e., that time was passing in which variations might have been occurring? (II) Again, couldn't value vary with temporal extent in this kind of time, A-time, even in the case of something which didn't itself vary during its period of existence? Consider the case of two four-dimensional

worlds, almost exactly alike and equivalently desirable. Let the situation be as follows: (1) Nothing exists; *then* (2) World 1 and World 2 exist; *then* (3) World 2 alone exists; *then* (4) Nothing exists. Couldn't the ethical profit gained through World 2's existence be greater just because it existed longer?

VIII. CONCLUSIONS

There is no hope of proving, by appeal either to experience or to logic, that the A-theory is right. All possible patterns of consciousness, of language-learning, of language-using, could be accounted for on the B-theory also, since any patterns logically able to exist one after another in A-time would seem also logically able to exist side by side along a dimension of a B-theory continuum. But facts like this cut both ways. As McTaggart's reasoning has such very doubtful success against the A-theory, we shall have to admit that neither logic nor experience can give conclusive proof of this theory's wrongness. True, some patterns in the physicist's records are more simply accounted for in a B-world, yet there could seem always to be the possibility that these patterns are in fact chopped up in A-time. And for a B-theorist to argue (as does Grünbaum¹³) that if there is non-relative Becoming then "the physicist must know it" is unfair, because—granted that, as just now said, any patterns existing in A-theory isolation could equally well exist in B-theory togetherness—the physicist could not possibly know it.

To the convinced axiarchist, however, the A-theory's logical and experimental viability will probably be of little interest. Even when he accepts the flow of A-time, he will be unlikely to think that the world flows with it; for how could a world which did so ever have come to exist? In a world existing in answer to an ethical requirement, wouldn't the A-theory's absolute changes alter matters only for the worse?

University of Guelph

Received April 4, 1975

¹³ *Op. cit.*, section 3.

III. ON SAYING

MARILYN FRYE

I

IN this paper I present a sorting and accounting of a variety of things which fall or might fall under the rubric "saying something." The object is clarification—the illumination of an area which can be a source of much confusion in discussion and analysis of speech acts. The point of departure is Austin's initial analysis of saying, in which he tries to set out the "acts" or "doings" which are supposed to be in some sense the elements of the total act of saying. From this, I develop an analysis of saying which may be seen either as an improved version of Austin's or as an alternative analysis, depending on one's assessment of the significance of its departure from Austin's initial intent or insight. I am inclined to see it as an alternative, though hardly a radical one.

To say something, according to Austin, is always to do at least three things: (1) to utter certain noises—a *phonetic* act; (2) to utter certain vocables or words, i.e., noises of certain types, belonging to and as belonging to, a certain vocabulary, conforming to and as conforming to a certain grammar—a *phatic* act; (3) to utter those words with a more or less definite sense and reference—a *rhetic* act. The act of saying something, analyzed into these three sub-acts, is then renamed by Austin the "locutionary act." He distinguishes the act of saying something, the locutionary act, from certain things we might be said to do in saying something—things such as asking or answering a question announcing a verdict, bequeathing, pronouncing sentence, warning, advising, suggesting. These latter he calls "illocutionary acts." The rhetic or locutionary act is, then that *in* the doing of which one is performing an illocutionary act.¹

This picture begins to run into trouble when Austin elaborates on his distinction between the phatic and rhetic acts. He says the following:

... "He said, 'The cat is on the mat'," reports a phatic act, whereas "He said that the cat was on the mat" reports a rhetic act. A similar contrast is illustrated by the pairs:

"He said 'I shall be there'," "He said he would be there";

"He said 'Get out'," "He told me to get out";

"He said 'Is it Oxford or Cambridge?'," "He asked whether it was in Oxford or Cambridge."

(p. 95)

Austin claims here that direct quotation is suitable for reporting phatic acts and indirect quotation is suitable for reporting rhetic acts. But this forces him to associate "He told me to . . .", "He asked . . ." and "He said that . . ." with rhetic acts. Telling and asking clearly are illocutionary acts, and "He said he would be there" should also report an illocutionary act, if the parallelism holds up. And thus it turns out that there are not as many reports as Austin thinks there are acts.²

(1) phonetic act

(1) He made some noises

(2) phatic act

(2) He said "The cat is on the mat."

(3) rhetic-locutionary act

(3) He said that the cat was on the mat

(4) illocutionary act

This lop-sidedness has been taken to show that there aren't as many differences here as Austin thought—that he made more distinctions than there are differences to distinguish them. Although I am not inclined to argue that there are as many different *acts* here as Austin thought there were, an analysis of saying reveals that

¹ J. L. Austin, *How To Do Things With Words* (New York, 1962), Lecture VIII. All parenthetical page references to Austin are to this work.

² This point has been noted by John Searle, "Austin on Locutionary and Illocutionary Acts," *The Philosophical Review*, vol. 77 (1968), pp. 405-424. Parenthetical page references to Searle are to this paper.

there are at least as many distinctions to be drawn as he thought there were. Not all distinctions in the analysis of saying need be distinctions among different kinds of acts; and anyway, outside of some context which provides relevant criteria for counting acts, one cannot say how many acts there are in a given situation.

Austin gives (2) as a report of a phatic act.

(2) He said "The cat is on the mat."

It seems to me that in ordinary discourse, a direct quotation such as this can be used to report the mere uttering of an utterance or to report something richer—a *saying*, which, unlike mere uttering, involves (at least) the speaker meaning something by the words and doing something on the order of referring.

That there is a sense of "say" such that "say" is roughly equivalent to "utter" is reasonably obvious. I can say Russian sentences whose meanings I do not know. If, on cue, I do this, I will still not have said anything, in one or more other senses of "say." If we want to mark this distinction with different reports, we can reserve the ordinary direct quotation format for the richer report, and invent another format for reporting the uttering of an utterance. For this purpose, I introduce (1') and call it a sentence-quotation report.

(1') He said 'The cat is on the mat.'

It might be objected that the "said" in ordinary direct quotation, as it is ordinarily used, is not ambiguous as between *utter* and some other sense of "say," but on the contrary, in that context, always means *utter*.³ In reply to this I want to point out that the phrases which can be substituted for "said" in

"The hammer is in the drawer," she said,

fall quite neatly into two classes. In the one group are "mumbled," "shouted," "roared," "intoned," "muttered" and so on, and in the other are "told him," "insisted," "warned," "answered," "added," "complained," "remarked" and so on. Of all these phrases, including "said," "said" itself is the only one which is at home in either class. And this is as solid a sort of case as one can have for the claim that a word, as occurring in a certain context, is ambiguous.

Using the sentence-quotation report for report-

ing utterings, we now have four different reports where Austin had three

(1) He made some noises.

(1') He said (uttered) 'The cat is on the mat.'

(2) He said, "The cat is on the mat."

(3) He said that the cat was on the mat.

But now, if the distinction between (1') and (2) is the distinction between merely uttering a sentence and saying something, then (2) and (3) both report sayings (in the richer sense[s]), and differ only uninterestingly, in that (3) purports to give only a paraphrase of what was said, while (2) purports to give exactly what was said. So, although we have four structurally different reports, we still don't seem to have four different things reported. Perhaps there are no more "things" to be reported—it depends on how we count "things." But the word "said" as it occurs in (2) and in (3) is still ambiguous, so there are still more different reports to be filtered out here, even if there are not more different things to be reported.

"Say" or "said" can be used in ordinary direct quotation whether the quoted sentence is indicative, imperative or interrogative.

She stood up and said, "Smith is a Fascist."

The stranger approached me and said, "Where's the police station?"

When Jones heard of her daughter's mischief, she turned to her and said, "Go to bed now, without your supper."

This neutrality of "say" is also apparent in the appropriateness of its use in all of the following little dialogues.

What did she say? She just asked me whether I wanted to go to the movies.

What did she say? She told me to return the book.

What did she say? She begged me to stay here.

Asking, suggesting, ordering, pleading are all covered by "say." When we report that someone said something, we may leave it quite open whether she suggested, reported, ordered, claimed, promised or so on. In a situation where we can see that a person is conversing (saying something, rather than merely uttering utterances), but cannot tell whether she is reporting or ordering,

³ I am here indebted to Kenneth Small.

questioning, protesting, greeting or betting, we may with propriety report, "She's saying something to the woman in the trenchcoat."

"Say," in some of its uses, serves as a sort of generic illocutionary verb whose specific terms include a wide range of illocutionary verbs which might be classed roughly as "assertive," "interrogative" and "imperative" as well as some, like "bet," "bequeath" and "bless" which perhaps do not fall into these classes.⁴

The "said" in the direct quotation report

(2) He said, "The cat is on the mat."

may be (though it need not be) read as this highly generic illocutionary verb. But in

(3) He said that the cat is on the mat

it seems that the presence of the nominalized indicative sentence restricts the range of "said" to what have been called "assertive" illocutionary acts. We might conjecture that there is another sense of "said" such that "said" is generic with respect just to the assertive illocutionary verbs, and then we might read (2) as ambiguous as between the higher level and the lower level generic illocutionary verbs.

In addition to all this, it seems to me that "say" is sometimes used in such a way that it must be construed as a garden variety specific assertive illocutionary verb in its own right. As such, it is a bit weaker than "state," and a bit firmer than "indicate," less defensive than "allege" and less formal than "assert." And when "said" occurs either in direct quotation contexts or in indirect quotation contexts, or even in contexts like "She said something to her," it may be this sense which is invoked.

There is one more sort of occurrence of "say" which deserves mention. In cases in which an illocutionary act is attempted but bogs down in infelicity,⁵ *something* has been done, nonetheless. And it seems that we should be able to use a direct-quotation report using "said" to report that act. If the use of this "said" is not to imply that some illocutionary act was actually performed, it should be in a way weaker than the "said" that I have characterized as the most generic illocutionary verb. On the other hand, to report what *was* done when an illocutionary act

did not come off, we seem to need a "said" that is stronger than "utter." The person who set out to threaten someone with a punch in the nose but failed through infelicity did not *merely* utter a sentence; she is sharply to be distinguished from a parrot. She uttered the utterance with the intention of performing some illocutionary act. And "say" is sometimes used to cover just such cases. When it is so used, it seems that it could be glossed as *utter an utterance with the intention of performing some illocutionary act*.

The picture so far is this. Where Austin had four different "acts," we have

- (A) making noises
- (B) uttering an utterance (a sentence)
uttering an utterance with the intention of performing some illocutionary act
- (C) performing some illocutionary act or other
performing some assertive illocutionary act or other
performing the particular assertive illocutionary act of saying something

All of the things under (B) and (C) can be covered in one context or another by the rubric "saying something." I have grouped them into groups (B) and (C) because it seems to me that the items in (B) would be classed with Austin's phatic act and the items in (C) fall under his heading of illocutionary acts.

(Certain readers will wonder what has happened to the rhetic-locutionary act. Observations pertinent to this question emerge in the next section.)

II

Although there has been a good deal of discussion about locutionary and illocutionary acts in the literature inspired by Austin, rather little attention has been paid to phonetic and phatic acts, which fall under the heading here of "merely uttering a sentence." There turn out to be some bugs lying under this rock; in particular, there is a distinction to be found that seems to map rather nicely on to Austin's distinction between the phatic and rhetic acts.

Characterizing the phatic act, Austin says it is "the uttering of certain vocables or words, i.e.,

⁴ As Annette Baier has reminded me, there really are enough illocutionary verbs which do not fall neatly into the "mood" classes to make this a significant simplification.

⁵ Austin, Lecture II, for discussion of infelicities.

noises of certain types, belonging to and as belonging to, a certain vocabulary, conforming to and as conforming to a certain grammar" (p. 95). This can be abbreviated to: the phatic act is the act of uttering a sentence *as a sentence*.

To understand what it is to utter a sentence as a sentence, it is useful and perhaps necessary to see what sorts of cases would be examples of uttering a sentence but *not as a sentence*. And since we are dealing primarily with cases of native speakers of the language in question who would not normally fail to know a sentence was a sentence, it would be a good thing to find a case of someone producing a sentence, recognizing it as a sentence, and yet not producing it *as a sentence*.⁶ With due apologies, I revert first to a case involving chess.

Suppose that an experienced chess-player approaches a chess-board which is all set up for a game. She reaches out and moves the White King's Knight from KN₁ to KB₃. She can't fail to know that this is an admissible move in chess. But suppose she is not interested in any aspect of the game of chess at the moment, but rather, she is interested in inspecting a small scratch on the surface of the board, on the square on which the Knight was standing. She moved the Knight merely to get it out of the way. If this is the case, one might plausibly say that her moving the Knight as she did was not something done *as a chess-move*. And the plausibility of this lies in the fact that her intentions, interests or wants, and the objects of her attention, had nothing to do with what chess moves have to do with. And, what is also important, her moving the piece just as she did, i.e., to KB₃, rather than anywhere else on or off the board, was completely irrelevant to the satisfaction of the intentions or

interests she had in changing the location of the piece. This suggests, for the case of the production of a sentence, that a person produces a sentence not as a sentence just in case its being a sentence is not relevant to the interests, purposes and so on that motivate the production of the sentence.

But now it must be noted that if a thing's being a sentence is relevant to the satisfaction of the interests, etc., which motivated its production, that relevance may yet be in virtue of any of the characteristic features or aspects of sentences. In the case of sentences produced by a student typist, for instance, it is relevant that what are produced be sentences, and this is so because sentences of a given language have certain physical features.⁷ But there are also cases in which the relevance in question is in virtue of the semantic properties of the sentence. The most obvious such cases would be those in which a speaker is making a statement or asking a question or the like—performing an illocutionary act. But there is another example in which no illocutionary act is performed. A person making up a multiple choice examination may produce a series of sentences which will appear as the items among which the student must choose. To satisfy the interests which motivated their production, the sentences must *mean* certain things and not others.⁸ A sentence is suitable for inclusion in the list of choices in virtue of its meaning, not in virtue of its physical properties.⁹

Apart from cases in which an illocutionary act is performed, there seem to be two sorts of cases of producing sentences as sentences—those in which the relevant features of the sentence are non-semantic features, and those in which the relevant features are semantic features. And this

⁶ Supposing what is produced is a sentence of a language is only supposing it is a string of sounds or of marks which, observed in isolation, would be taken by a native speaker of that language to be a sentence of that language. Thus, in principle, the wind could produce a sentence in the sand. Neither a human, nor even a sentient agent is logically required for the production of a sentence.

⁷ Although these physical features are associated in systematic ways with semantic properties, the semantic properties are not at all the properties in virtue of which it matters that the things produced be sentences. That this is so is shown by the fact that sentences can be replaced in typing exercises by non-sentences that preserve the relevant physical properties, or most of them. This dispensability of the sentences does not, however, show that their being sentences is not after all appropriately relevant. Even where their being sentences is not *necessary* to the satisfaction of the interests which motivated their production, there still are features of sentences which make them appropriate for production in these circumstances, and that is still at least part of the reason they were produced. The "relevance" of which I speak is nowhere to be assumed to be a necessity.

⁸ The interests in question do not necessarily uniquely determine what meaning a sentence must have in order to be suitable, but they determine it within quite narrow range.

⁹ Though it seems clear to me that no illocutionary act is intended in this case, it may be even clearer if we imagine the teacher to be merely practicing making up multiple choice examinations. (This revision of the example is due to the audience who responded to an earlier version of this paper delivered at the University of Washington.)

division, it turns out, fits rather well with Austin's characterizations of the phatic and rhetic acts. The phatic act is supposed to be the act of uttering certain vocables belonging to and as belonging to a certain vocabulary, conforming to and as conforming to a certain grammar, i.e., uttering a sentence as a sentence. And the rhetic act is supposed to be the act of "using those vocables with a certain more-or-less definite sense and reference." If the sentence produced is, in virtue of its meaning, suitable to the satisfaction of the interests which motivated its production, it seems not wholly inappropriate to say that it was not only produced as a sentence, but was produced *as having a certain more-or-less definite meaning*.

Since I am not inclined to accept the view that there is a separate *kind of act* picked out by Austin's term "rhetic act," I would not recommend the use of that term here. But I believe it is clear enough now that among instances of producing sentences, there are instances of their not being produced as sentences (phatic), and instances of their being produced as sentences, and that among the cases of their being produced as sentences, there are instances where the

meanings of the sentences are not pertinent (phatic), and instances where they are (rhetic).¹⁰

The picture of the speech act that has emerged here may be outlined as follows: Sounds are made (phonetic), which may or may not constitute a sentence; where the sounds constitute a sentence, that sentence may or may not have been produced as a sentence; where the sentence was produced as a sentence (phatic), the relevant features of the sentence, which suit it to the interests which motivated its production, may or may not be semantic features; where the sentence is suitable to the satisfaction of these interests in virtue of its meaning (rhetic), it may or may not have been produced with the intention of performing some illocutionary act; and when a sentence is produced with the intention of performing some illocutionary act, an illocutionary act may or may not actually have been performed. The verb "say" in one or another of its meanings or senses is applicable to both sorts of cases of the production of a sentence as a sentence, to the production of a sentence with the intention of performing an illocutionary act, and, with three degrees of generality, to the performance of illocutionary acts.¹¹

Michigan State University

Received June 24, 1975

¹⁰ I believe the example of the exam-preparer shows that there are cases of this last sort which yet are not cases of the performance of an illocutionary act. In the dubious vocabulary of "acts", this case might constitute a counter-example to the claim that all locutionary acts are illocutionary acts (cf., Searle, *op. cit.*, p. 413). Whether it really does or not may depend on whether we must assume the producer of the exam-sentences means something by those sentences. For the point about locutionary acts seems to be connected with the claim that locutionary acts should be cases in which the utterer means something by the utterance, and all cases where the utterer means something by the utterance are claimed to be cases of illocutionary acts or attempted illocutionary acts. My guess is that the meaning of the expression "mean something by the utterance" is not precise enough to yield a definite verdict about whether the exam-preparer meant something by her utterance.

¹¹ I am indebted to the critical comments of Steven Davis, Murray Kiteley, Christopher McMahon and Kathleen Owens. I also wish to acknowledge that my views on these topics developed from things I first learned from Paul Ziff.

IV. NEWCOMB'S PROBLEM: FURTHER INVESTIGATIONS

DORIS OLIN

THE powerful and intriguing paradox discovered by Dr. William Newcomb concerns a game played with a Being who has extraordinary predictive powers. The game can be described as follows. There are two boxes, B_1 and B_2 . B_1 contains a thousand dollars. B_2 contains either a million dollars or nothing. You have a choice between two actions: taking what is in both boxes or taking only what is in B_2 . Before you act, the Being examines you, and on the basis of his findings, predicts what you will do. If he predicts that you will take only what is in B_2 , he places a million dollars in B_2 ; but if he predicts that you will take both boxes, he leaves B_2 empty. (You are not allowed to randomize your choice, by for example, flipping a coin.) Once the Being has acted, it is your turn. You are fully informed concerning the rules of the game; and since you know that the Being has successfully predicted the choices of all previous players, you believe it almost certain that he has correctly predicted your choice. What should you do?

What makes the Newcomb problem so disturbing is the fact that, for each alternative, a strong and persuasive case can be made that it is the rational choice. Briefly, the reason for preferring the two box strategy is that you play after the Being and hence whether there is anything in the second box is already determined by the time you take your turn. So whatever the Being did, you stand to gain \$1,000 more by taking both boxes rather than just B_2 . But the counter argument for choosing just B_2 is that given the Being's incredible record of correct predictions, it is almost certain that the one box strategy will yield a prize of \$1,000,000; whereas, almost certainly, taking both boxes will yield only \$1,000.

The Newcomb problem was first presented and set in the context of contemporary decision theory by Robert Nozick.¹ In particular, Nozick shows that the Newcomb case raises a serious problem for

decision theory. For in this situation, each action seems to be recommended by different principles of choice; and it is by no means clear which principle should be followed. First, there is the expected utility principle which states that you should act so as to maximize expected utility. If the (exclusive and exhaustive) outcomes an action A may yield are $O_1, \dots, O_n$, the probability of O_i given that A is done is p_i , and the utility of O_i is $u(O_i)$, then the expected utility of A is $p_1 \times u(O_1) + \dots + p_n \times u(O_n)$. Assuming that the probability of the Being's predicting correctly is only .9, the expected utility of taking only B_2 is $.9 \times u(\$1,000,000) + .1 \times u(\$0)$, while the expected utility of taking both boxes is $.1 \times u(\$1,001,000) + .9 \times u(\$1,000)$. No doubt the utility assigned by the reader to the different outcomes is such that taking only the second box has a higher expected utility than taking both, and is therefore the action recommended by the expected utility principle.

But a case can be made for taking both boxes by appealing to another, equally respected, principle of choice. Note first that the game can be represented by the payoff matrix below.

	S_1	S_2
A	\$1,001,000	\$1,000
B	\$1,000,000	\$0

where: S_1 = The Being placed a million dollars in the second box.

S_2 = The Being left the second box empty.

A = You take both boxes

B = You take only the second box.

The second principle of choice, dominance, can be stated as follows. Let a_i be the utility entry in a payoff matrix, in the cell for action A and state S_i . If the payoff matrix for actions A and B and a partition of states of the world $S_1, \dots, S_n$ is such that for every i , $a_i > b_i$ and for some i , $a_i > b_i$, then A

¹ Robert Nozick, "Newcomb's Problem and Two Principles of Choice" in *Essays in Honor of Carl G. Hempel*, ed. by Nicholas Rescher (New York, 1970), pp. 114-146.

weakly dominates B relative to that partition. The dominance principle says that if A weakly dominates B relative to some partition, then A ought to be preferred to B .

Looking again at the matrix above, and making reasonable assumptions about the utility assigned to the different payoffs, we can see that in the Newcomb case, A is the dominant action. Indeed, A strongly dominates B (i.e. for every i , $a_i > b_i$).

Faced with a situation in which two principles of decision theory seem to conflict, Nozick attempts to determine the conditions for the legitimate application of the dominance principle. Richard Jeffrey, who considers this problem outside the context of the Newcomb paradox, claims that appeal to the principle is warranted only when the states are probabilistically independent of the actions—a position which Bar-Hillel and Margalit continue to maintain in their article on the Newcomb paradox.² Nozick, however, after a lengthy discussion, concludes that when the actions do not causally influence which state obtains, one should perform the dominant action. Each supports his case only by giving examples where the dominance supposedly does (or does not) apply. One serious danger of such a procedure, of course, is that one may fail to isolate the *relevant* features of the examples studied. A more direct approach, if possible, would surely be preferable.

The Newcomb case is thus relevant to formal decision theory in that it raises, in a particularly forceful way, the problem of what sort of constraints should be placed on the application of the dominance principle; and in the light of this case, neither of the two theories advanced seems completely obvious and intuitive. I shall try to resolve this problem in a way which does not rest on our intuitions about particular cases, and which also enables us to understand *why* it is fallacious, in certain circumstances, to argue that an act ought to be done because of dominance considerations. But, the reader should be forewarned, this paper will not issue in a proof that a particular strategy should be followed in the Newcomb case. Rather, my main purpose, with regard to the Newcomb problem, is to bring the key issues into clearer focus and to determine the questions which must be answered if the problem is to be satisfactorily resolved. I shall first establish the conditions for

the legitimate appeal to dominance considerations and then show that having this information does not automatically enable us to dispose of the Newcomb case (although it does make it much easier to cut through to the crucial issues). A complete resolution of the Newcomb problem requires an answer to another question which does not fall within the scope of rational decision theory.

The best approach to the problem of determining the relevant conditions is to consider why it is that if the dominance principle is correctly applied and recommends A , this provides good reason for doing A . Restricting attention for the moment to cases in which A strongly dominates B , one could point out the following. First, supposing that I did A and got x units of utility, I would then be assured that if I had done B , I would have gotten less than x . So there will never be room for regret if I do A . More generally, prior to acting, I could say "No matter what state obtains (obtained, will obtain), if I were to do A , I would be better off than if I were to do B ." Such a statement must be true if application of the dominance principle is justified; and if it is true, it seems clear that I ought to do A .

Suppose now that the matrix for a certain situation is as below.

	S_1	S_2
A	10	4
B	8	2

To establish the truth of "No matter what state obtains, if I were to do A , I would be better off than if I were to do B ," I would have to argue as follows.

- (1) Either S_1 obtains or S_2 obtains.
- (2) Suppose S_1 obtains.
- (3) If I were to do A , I would get 10.
- (4) If I were to do B , I would get 8.
- (5) Suppose S_2 obtains.
- (6) If I were to do A , I would get 4.
- (7) If I were to do B , I would get 2.
- (8) Therefore, no matter what state obtains, if I were to do A , I would be better off than if I were to do B .

What is crucial, and might be overlooked is that (3) and (4) must be true on the supposition of S_1 , and (6) and (7) must be true on the supposition of S_2 . But having recognized this, it is now

² Richard Jeffrey, *The Logic of Decision* (New York, 1965). Maya Bar-Hillel and Avishai Margalit, "Newcomb's Paradox Revisited," *British Journal for the Philosophy of Science*, vol. 23 (1972), pp. 295-304. A state S is probabilistically independent of actions A and B if the probability of S given that A is done equals the probability of S given that B is done.

easy to see why there should be general agreement on the following: The dominance principle cannot be legitimately applied to this case if doing A would cause S_2 and doing B would cause S_1 . Note first that the appropriate tense for (1), (2) and (5) may vary, depending on the details of the particular case. If the actions cause the states, then (while considering which action to perform), we should use the future tense. (I assume throughout that backwards causality is impossible.) But if doing A would cause S_2 , then we cannot argue as follows:

- (2) Suppose S_1 will obtain.
- (3) If I were to do A , I would get 10.

For if A would cause S_2 , then even though S_1 will in fact obtain, the following is still true: If I were to do A , S_2 would be the case and I would get 4.³ But this last conditional is clearly incompatible with (3). (The two are contraries.) So we cannot say that (3) is true on the assumption of S_1 , and the dominance argument is therefore defeated.

To return to the question with which we began, let us try to restate the dominance principle so as to incorporate the necessary conditions.

Dominance Principle: Let $a_1, \dots, a_n$ and $g_1, \dots, g_n$ be utility numbers. Suppose there is a partition of states of the world $S_1, \dots, S_n$ such that for every i , on the assumption that S_i obtains (obtained, will obtain) it is true that (1) If I were to do A , I would get a_i and (2) If I were to do B , I would get b_i . Then if for every i , $a_i \geq b_i$ and for some i , $a_i > b_i$, A should be preferred to B .

(To say that "If I were to do A , I would get a_i " is true on the assumption that S_i obtains is to say that if " S_i obtains" were true, then "If I were to do A , I would get a_i " would also be true.)

To contrast this formulation of the principle with the previous one, note that a payoff matrix provides information concerning the result of the joint occurrence of a certain action and a certain state. Thus the matrix just considered tells us that the result of the occurrence of both S_1 and A would be 10 units of utility. But from this alone we cannot infer that if " S_1 will obtain" were true, then "If I were to do A , I would get 10." would be true. A slight variation of an earlier example may help to

make it clearer that this is not a valid pattern of inference. The result of the joint occurrence of living 20 years after t and jumping off the Empire State Building at t (with no protective devices) would be, say, to become the object of intense medical interest. But even though "I will live 20 years after t " is true, we would not want to grant "If I were to jump off the Empire State Building at t (with no protective devices), I would become the object of intense medical interest."

Thus the general problem in decision theory raised by the Newcomb case is now resolved. What has emerged is that the relevant conditions for the legitimate application of the dominance principle are just those necessary to establish the truth of a complex comparative proposition. Having restated the principle so as to incorporate these conditions, we can now say simply that if it recommends act A , one ought to do A . (Similar requirements should of course be built into the maximin principle; although, since the maximin strategy is too conservative for certain situations, we still would not wish to advocate its unrestricted use.)

To return to the Newcomb problem, the question remains whether the dominance principle can be applied to this case. Certainly, the argument below looks impeccable at first sight.

- I. (1) Either the Being placed \$1,000,000 in B_2 or he left it empty.
- (2) Suppose the Being placed \$1,000,000 in B_2 .
- (3) If I were to do A (take both boxes), I would get \$1,001,000.
- (4) If I were to do B (take just one box), I would get \$1,000,000.
- (5) Suppose the Being left B_2 empty.
- (6) If I were to do A , I would get \$1,000.
- (7) If I were to do B , I would get \$0.
- (8) Therefore, no matter what state obtained, if I were to do A , I would be better off than if I were to do B .

But the advocate of the one box strategy will claim he has excellent evidence for the premises of the argument below.

- II. (1') If I were to do A , the Being would have predicted this, would have left B_2 empty, and I would get only \$1,000.

³ An example may help here. Suppose that the result of my jumping off the Empire State Building at t (with no protective devices) and being dead shortly after t would be -10 units of utility. Then even though I will in fact be alive long after t , the following is still true: If I were to jump off the Empire State Building at t (with no protective devices), I would be dead shortly after t and I would get -10.

- (2') If I were to do *B*, the Being would have predicted this, would have put \$1,000,000 in *B*, and I would get \$1,000,000.
- (3') Therefore, if I were to do *B*, I would be better off than if I were to do *A*.

Let us try to see more clearly why the two arguments yield incompatible conclusions. In order for argument I to be correct, (2) and (4) must be true on the assumption of (2), and (6) and (7) must be true on the assumption of (5). If this is in fact the case, then since either (2) or (6) is true, either (3) and (4) are true or (6) and (7) are true. But the conjunction of (3) and (4) is incompatible with (1'), for (3) and (1') are contraries; similarly, the conjunction of (6) and (7) is incompatible with (2'). (If we preface the premises and conclusion of II with "It is almost certain that . . .," the two arguments will no longer conflict in this way. But we can still say that one cannot reasonably accept both. For if *p* and *q* are incompatible, it cannot be reasonable to believe that *p* is true *and* that it is almost certain that *q* is true. Alternatively, if we rewrite (1') and (2') as material conditionals, there will be no conflict with argument I; but then we cannot conclude, from these new premises, anything which provides a reason for preferring *B*. To see this, we need only note that it is sufficient for the truth of the corresponding material conditionals that I will in fact do *A*, get \$1,000, *etc.* And these facts are quite compatible with "If I were to do *B*, I would get \$0.")

It is, I think, the plausibility of both I and II which serves to create the conflict in the Newcomb case. Of course, neither argument represents anything novel to discussions of the Newcomb problem. In fact, each is a (more explicit) version of the informal arguments standardly offered on behalf of each strategy. But what has never been noticed (perhaps because the conclusion of II is usually left implicit), is that the empirical claims made in the two arguments are not all compatible and cannot all be maintained. As a result of this oversight, there has been a natural tendency to assume that the problem posed by the Newcomb example is that of deciding which of two sets of facts provides a better reason for performing an action.

Having seen that the arguments for each side make conflicting presuppositions concerning the empirical facts, there is no longer any difficulty in understanding how the argument for one strategy, if correct, would destroy the force of the other; and it is now also possible to isolate one of the key issues

involved in determining which strategy should be preferred. In order to provide a complete resolution of the Newcomb problem, one question which must be coped with is: Given the description of the Newcomb case, is it more reasonable to believe (1') and (2') or to suppose that (3) and (4) are true on the assumption of (2); *etc.*? An attempt to grapple with this question should take into account the following observations.

Whether (1') and (2') ought to be accepted depends entirely on the credibility of

- p*: If I were to do *A*, the Being would have predicted that I would do *A*.
- q*: If I were to do *B*, the Being would have predicted that I would do *B*.

Once convinced of the truth of these conditionals, there should be, given the rules of the game, no further hesitation in accepting (1') and (2'). Lest the appropriate judgment seem patently obvious to some readers at this point, it must be noted, first, that even if we grant "The Being knows what I will do," this does not guarantee the truth of *p* and *q*. For suppose Smith predicts that, when given a choice between tea and coffee at lunch today, I will choose tea. His prediction is based on the fact that I have always chosen tea over coffee in the past. Given that his belief is true and based on the right kind of evidence, we should admit that he knows what I will do. But granting this fact surely need not prevent us from saying, at the moment of choice, that if I were to take coffee, Smith would have made a false prediction.

Still, it is possible to flesh out the Newcomb case, in particular to specify the *nature of the evidence* on which the Being bases his predictions, in such a way that *p* and *q* are both true. If the Being finds, as a result of his examination, that you are in state *C*₁ one hour before the time you must play, he predicts that you will do *A*; if he finds that you are in *C*₂, he predicts *B*. Further, the states used as evidence are nomically linked to the actions. There is, let us suppose, a law which says that all and only those who do *A* are in state *C*₁ one hour before playing; similarly with regard to action *B* and state *C*₂. This will ensure the truth of

- r*: If I were to do *A*, I would have been in the state *C*₁ one hour before playing.

And there is no difficulty in granting the truth of

- s*: If I were to do *A* and to have been in state *C*₁ one hour before playing, the Being would have predicted that I would do *A*.

Together, these conditionals guarantee the truth of *p*. (If we wish to make it part of the Newcomb case that the Being occasionally predicts incorrectly, we could add that very, very infrequently the Being is mistaken about which state the subject is in — *e.g.* because his instruments malfunction.)

Having said as much as I can confidently say, let me conclude by indicating some areas of uncertainty. I have argued that the conflict in the Newcomb situation rests just on two conflicting views of the empirical facts. But it is by no means evident that the philosophical interest of the problem is now exhausted. For one thing, even if argument II is sound, it is still not clear that this automatically guarantees that *B* is the rational act. For suppose (1') and (2') are true because the evidence on which the Being bases his predictions is as just described. In such a case, many philosophers would say that the agent is not free to do either *A* or *B*, that there can be only one strategy open to him. But if the conditions which yield the truth of (1') and (2') also conflict with the assumption that the player can do either *A* or *B*, then we are faced with yet another serious problem, *viz*: Given that there is only one strategy open to the agent, is it legitimate to characterize that strategy as either rational or irrational? Certainly there is considerable temptation to think that the current orthodoxy concerning moral appraisals and freedom should be extended to judgments of the rationality of actions.

Further, quite apart from any inferences one might wish to draw concerning rationality, the question of which argument is best supported by the evidence may well be linked with larger philosophical issues. In order to make a case for argument I, we must deny that the Being bases his

predictions on past states which are nomically related to the actions in the way I have described. But given the overwhelming evidence of the Being's complete success to date, it is hard to see how we could plausibly justify this denial without resorting to one of the following. (1) We might argue that it is conceptually impossible that the choice one makes, in deciding between two actions, should be nomically determined by past states. (2) Or, we might try to show that the supposition that the choice one makes is thus determined conflicts with certain other assumptions so fundamental or well-established that not even a correlation of the sort envisioned would warrant their rejection. At the moment, I can see no plausible way of blocking the inference to the relevant laws without endorsing one of these positions concerning choice in general. Thus, if the reader cannot escape the uneasy feeling that *A* must be the rational act, he should perhaps begin to question whether a deterministic view of choice can be meshed with our present conceptual framework as easily as many of us would like to think.

What has been established is that, assuming the player is in fact to do either *A* or *B*, the question of which strategy should be preferred is no longer particularly a problem for decision theory. For the crucial issue which now confronts us in determining which alternative is rational is not the legitimate scope of the dominance principle; nor does the dispute center on whether one set of empirical facts provides a better reason for performing an act than does another. Rather, the problem is simply to decide what the empirical facts are (alternatively what, given the evidence, it is reasonable to believe them to be).⁴

Glendon College, York University

Received January 10, 1975

⁴ I wish to thank my colleague, Professor Nollaig MacKenzie, for helpful discussions of this problem.

V. NOTES ON THE LOGIC OF THE IGNORANCE RELATIONS

SIDNEY AXINN AND DAVID AXINN

THIS paper outlines an elementary formal logic of the ignorance relation. Various epistemic logics have been developed and they permit some analysis of ignorance, but they don't allow for several important distinctions. Ours is a semantic conception of ignorance, and, like Tarski's classic paper on truth, we make special use of the distinction between use and mention. We separate three kinds of ignorance: 1) inability to *mention* a sentence, 2) inability to assert *the truth-value* of a sentence, and 3) the inability to assert a sentence that correctly makes a *reference*.

For an individual to be ignorant, it must be the case that some sentence is true, and our individual is not capable of asserting it when examined. Ignorance can be distinguished from *abnornance*: a case of ignorance occurs when there is something accurately assertable, abnornance occurs in the absence of something properly assertable. If we can't say what our weight is, we are ignorant of the matter. If we can't say what color wings we have, we are abnorant.

Section #1 offers definitions and section #2 compares them with other modal categories. Section #3 gives some axioms that are partly presented in section #4 on a square of opposition. Section #5 makes further assumptions and considers some of the formal results. Roman numerals are used for the axioms, arabic numerals for the assumptions, and a 'T' and number precede the statements for which we argue. Ignorance and abnornance are symbolized by their first consonants.

1. Definitions

We will take individuals to be things capable of making a statement in some language, without regard to whether or not they happen to be humans. The act of making a statement may consist of uttering a sentence, or of any other form of communication; pointing, writing, circling, typing,

etc. We assume the existence of some examiner, possibly but not necessarily the same as the subject. The ignorance relations we will develop are a means for the examiner to specify the extent of the subject's ignorance.

We take ignorance to be ignorance of a sentence or set of sentences: ignorance of things, propositions or statements can be expressed by well-chosen sentences about them. The utterance of a sentence is called a statement, but utter-failure still leaves us with an object of the ignorance relation; a sentence need not be uttered, as Quine has argued.¹ Ignorance of a set of sentences will be taken as a convenient way of describing ignorance of each member of the set. In order to specify the truth of a statement asserting ignorance, we must also specify the time at which it is uttered. The system of real numbers will be used to name the time values, although we will often omit the times.

1.1. Ignorance as an Image Relation

The image of a class C by a relation R , the R -image of C , is presented in Quine's formulation as the class of all those elements which bear R to one or more members of C .² Where R is the teacher-relation and C the class of mathematicians, the R -image of C , $R"C$, indicates the class of teachers of mathematicians. Following his definition #28 (p. 210), the $R"C$ would be an abbreviation for $\hat{x}(\exists y)(R_{x,y} \ \& \ y \in C)$ — those who are teachers of someone who is a member of the class of mathematicians. Let us replace R by G , and C by L , and take the class L to be a particular class of sentences: then the G -image of L will specify those who are ignorant of sentences that are members of the class L . In Quine's definition (and examples) classes are used to specify both places in R , but individuals or their unit classes will work quite smoothly. An individual can then be said to bear the relation of ignorance to a particular sentence that is a member

¹ W. V. O. Quine, *Word and Object* (Cambridge, 1960), p. 194.

² W. V. O. Quine, *Mathematical Logic*, Revised Edition (Cambridge, 1951), pp. 209-214. Unless indicated, further Quine references are to this work.

of a certain class of sentences. This last class of sentences may be the unit class of the indicated sentence. We could have obtained the different assertions of ignorance displayed (below) in Table I by changes in the sentences that occupy the second place of the G -relation, rather than by changes in the relation itself.

By constructing a definition of ignorance as an image relation, we get a device for asserting ignorance of either a particular sentence, a part of a set of sentences, or a whole set of them. Of course, following Quine's definition of image, for the ignorance relation to hold there must exist something that is the sentence of which our subject is ignorant.

There are two special cases that are of interest — ignorance of the truth value of a sentence, and ignorance of the fact that something is a sentence. Each of these relationships will be expressed as ignorance of the set of sentences expressing these facts. Knowledge will be taken to be the negation of ignorance. Reading $G(a, s, t)$ as a is ignorant of s at t , we have:

Definition 1. $G(a, s, t) =_{df}$ a is unable to assert s at t , given that s is true. (a an individual, s , a sentence, and t a time-value).

Definition 2. $G(a, \{s\}, t) =_{df} \bigwedge_{s \in \{s\}} G(a, s, t)$, (where $\{s\}$ is a set of sentences, & $\bigwedge$ is the conjunction of all of its members).

Definition 3. $G^*(j, s, t) =_{df} G(j, \{s'\}, t)$, where $\{s'\}$ is the set of all sentences correctly expressing the fact that s is a sentence. Symbolically,

Definition 4. $G^*(j, \{s\}, t) =_{df} \bigwedge_{s \in \{s\}} G^*(j, s, t)$.
(Read: j is ignorant of the mention of $\{s\}$ at t).

Definition 5. $G^v(j, s, t) =_{df} G(j, \{s'\}, t)$, where $\{s'\}$ is the set of all sentences correctly expressing the truth value of s . Symbolically,

Definition 6. $G^v(j, \{s\}, t) =_{df} \bigwedge_{s \in \{s\}} G^v(j, s, t)$.
(Read: j is ignorant of the truth value of $\{s\}$ at t).

Definition 7. $K(j, s, t) =_{df} \neg G(j, s, t) \ \& \ s$. (Read: j knows s at t)

Definition 8. $K(j, \{s\}, t) =_{df} \neg G(j, \{s\}, t) \ \& \ \{s\}$.

Definition 9. $B(a, s, t) =_{df}$ a is unable to assert s at t , and s does not exist in his language. (Read: a is abnorant of s at t .)

One property of the ignorance of sets, defined as above, lies in the fact that ignorance of a set (Def. #2) requires ignorance of all of its members, but knowledge of it (Def. #8) requires knowledge of only one of the members.

We have already indicated a departure from the ordinary image relation. The G relation is to be a three-place rather than a two-place predicate, the third place to be a name for a time. And, as mentioned, the third place will often be tacit in this version. Some of Quine's theorems about images will serve us as support for developments of the G relations in lines T(13) through T(16) below.

Just what restrictions are there on occupants of the first two places of G ? The first place, ordinarily, is to be the name of an individual; and an individual has been defined above as a source of assertions. Of course such an individual may be a composite entity, a computer, a non-human animal, a human, etc. The sentence that fills the second place in G is to be a declarative. However, as long as it is taken to be a sentence in the subject's language, it may declare anything, e.g., a truth value, as in Definition #5, or declare that an ignorance relation holds in a certain situation, etc. Iteration is available, and of course it is significant. One can be ignorant of his own or someone else's ignorance. Some of this will be developed in a later section. Defining ignorance as an image allows us to express the cases in which an individual is ignorant of anything from one sentence to a whole language; as well as those cases in which many people are ignorant of either single sentences, several, or whole languages.

2. The Other Modal Categories

Von Wright takes modalities as "dyadic predicates or 'relations' between propositions."³ Whether ignorance is a modal operator is not perfectly obvious. However, this paper assumes it to be partly modal, and considers it to be a three-place relation holding between a person, a sentence, and a time. Ignorance is an *image* relation, as presented earlier.

It may be helpful to compare and contrast the distinctions between kinds of ignorance with the

³ "A New System of Modal Logic," *Logical Studies* (London, 1957), p. 89.

TABLE I. MODAL CATEGORIES

	Alethic	Existential	Epistemic	Deontic	Counter-epistemic	
a.	necessary	universal	verified	obligatory	unmentioned	$G^*(j, s, t)$
b.	possible	existing	unfalsified	permitted	unknown	$G(j, s, t)$
c.	contingent	partial	undecided	indifferent	undecided	$G^v(j, s, t)$
d.	impossible	empty	falsified	forbidden	known	$K(j, s, t)$

modal categories that are already well-known. Table I is based on von Wright, but taken from Anderson and Moore,⁴ except for the counter-epistemic column.

Anderson and Moore (and von Wright) stress some interesting properties that are displayed in Table I. Where the letters refer to entries in the table: "whatever is *a* is *b*, whatever is *b* is not *d*, and whatever is *c* is neither *a* nor *d*."⁵ The first two of these formal similarities also hold for the counter-epistemic mode. Some additional similarities also hold in each of the modes except the counter-epistemic, e.g., whatever is *b* is either *c* or *a*. They take the epistemic and alethic modes to apply to propositions. Presumably, the counter-epistemic mode could also apply to propositions, if wanted, with statements as their names.

3. Axioms

The distinction between the examiner's and the subject's language must always be kept in mind, although the two may overlap (e.g., when an individual makes statements about his own ignorance). The ignorance relations are contained in the examiner's language, and are used to make statements about the subject's ignorance of sentences in his language (it is assumed that the examiner has made an assumption specifying the language in which the subject is speaking). Names for the subject-individuals, sentences, and time values must be included in the examiner's language. We will assume a two-valued universe of discourse for both languages.

The following axioms present the minimal properties of the ignorance relation—further assumptions will be made in section 5. We assume that the following universal quantifiers of *x*, *y* and *z* are restricted by the conditions on the definition of

G: *x* must be an individual, *y* a true sentence, and *z* a time-value.

Axiom I, asserts that if an individual cannot even mention a sentence, then he can hardly assert its truth value and its reference.

A I. $(\forall x) (\forall y) (\forall z) [G^*(x, y, z) \supset (G^v(x, y, z) \& G(x, y, z))]$.

The second axiom holds that if someone is unable to assert the truth-value of a sentence, he cannot correctly use it.

A II. $(\forall x) (\forall y) (\forall z) [G^v(x, y, z) \supset G(x, y, z)]$.

Thirdly, if someone is ignorant of something, that thing must exist. Semantically, if one is unable to assert *s* when it is true, *s* is true.

A III. $(\forall x) (\forall y) (\forall z) [G(x, y, z) \supset y]$.

Definitions 2, 4, 6, and 8 may be used to derive axioms for ignorance of sets of sentences:

A IV. $(\forall x) (\forall y) (\forall z) [G^*(x, \{y\}, z) \supset (G(x, \{y\}, z) \& G^v(x, \{y\}, z))]$

A V. $(\forall x) (\forall y) (\forall z) [G^v(x, \{y\}, z) \supset G(x, \{y\}, z)]$

A VI. $(\forall x) (\forall y) (\forall z) [G(x, \{y\}, z) \supset (y_1 \& y_2 \& \dots)]$

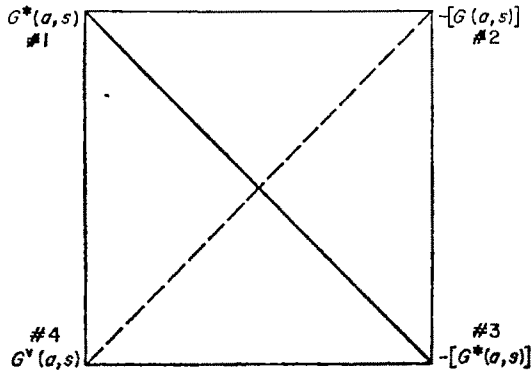
4. Square of Opposition for Ignorance

Accepting Axioms I and II, we find that the properties of the three ignorance relations fall into a partial square of opposition.

All but one property of a classical square of opposition are exhibited here. The diagonals seem to connect contradictories, but only one does. #1 and #3 are clearly contradictory; they differ only by a sentential negation sign. #2 and #4 cannot both be true: if one is ignorant of the truth value of *s* (#4), he must also be ignorant of *s*, according to Axiom II. However, they can both be false. One might know the truth value of the sentence *s*,

⁴ Alan R. Anderson and O. K. Moore, "The Formal Analysis of Normative Concepts," reprinted in J. M. Copi and J. A. Gould, *Contemporary Readings in Logical Theory* (New York, 1967), pp. 319-31.

⁵ *Ibid.*, p. 326.

TABLE II: *Partial Square of Opposition for Ignorance*

denying #4; but be ignorant of the reference of the sentence, denying #2. Consider a case in which one might know that a particular sentence in the *Handbook of Physics and Chemistry* is true (perhaps by its equivalence to a known true sentence, etc.) but not know when it may be asserted correctly. So, contradiction fails for this pair, and we indicate this by a dotted line. #2 and #4 are merely contraries. #1 and #2 should be and are contraries, as can be shown from Axiom 1. #3 and #4 should be and are subcontraries, again following from Axiom 1. The verticals, #2—#3, and #1—#4, connect subalterns, as Axiom 1 can establish.

In summary, the arrangement in Table II makes a traditional square of opposition with just the one failure that we have mentioned. Of course, other arrangements of opposition can be offered. Replacing the lower left entry by $G(a, s)$ gives a complete square, but doesn't contain each kind of ignorance.

5. Assumptions and Some Theorems

Now we can summarize some details and present things more systematically. We have already introduced our definitions and axioms. It may help to collect the assumptions that were involved. They are:

- A 1 *Ignorance* is established by the inability to correctly state a certain sentence or set of sentences.
- A 2 Knowledge of ignorance is relative to the *examiner*. Self-examination may occur.
- A 3 Following the usual convention, a sentence without quotes is *used* to refer to something other than itself, and a sentence within quotes refers to itself.
- A 4 There is just one language in the universe of discourse and it's two-valued.

- A 5 Any *individual* can be the subject in an ignorance relation if it is an entity that can communicate the assertion of a sentence, i.e., make a statement. We assume a responsible ignoramus, as defined below.
- A 6 The subject's language is the set of all sentences that he can assert. The examiner's language includes names for the sentences in the subject's language.
- A 7 A *correct assertion* of a sentence is one which satisfies the examiner. Uttering may be one form of mentioning: there are others such as pointing, writing, circling, etc.
- A 8 The ignorance relation is contained in the examiner's language.

Since we assume a two-valued language, for any true statement s and any individual a ,

$$T(1) \quad G(a, s) \vee -[G(a, s)].$$

The knowledge relation has been defined as the denial of ignorance (definitions #7 and #8), so

$$T(2) \quad K(a, s) \supset [s \ \& \ -[G(a, s)]]:$$

The knowledge relation has the same domains and conditions on its variables. (2) is what we would expect, in common with the other modal categories in Table I. Axiom I gives us

$$T(3) \quad G^*(a, s) \supset G(a, s)$$

and

$$T(4) \quad G^*(a, s) \supset G^v(a, s).$$

By transportation (3) gives us

$$T(5) \quad -[G(a, s)] \supset -[G^*(a, s)],$$

which rules out unspeakable truth.

We can no longer postpone a consideration of a problem that has been involved in studies of epistemic logic, the matter of opaque contexts. This requires us to specify the nature of rational, responsible or defensible ignorance. Is one ignorant of the logical consequences of his ignorance? Suppose that

$$(A) \quad P \equiv Q,$$

$$(B) \quad P \supset R,$$

and

$$(C) \quad G^v(a, P).$$

Does a 's ignorance of the truth value of P carry with it the assurance that he is ignorant of Q and of R ? He may not know (A) and (B). In either case, we must make some assumption about the rational powers or rational ability of the ignorant individual. The assumption we choose to make depends on the

use to which we will put the resulting logical machinery. For the general illustrative purposes of this paper, we will make the special-case assumption that we must distinguish between what is available to the subject whose ignorance is being specified and what is available to the constructors of the system of assertions about the individual. The constructors of the system, we outside observers, are taken to have all the machinery of logic and the truths of science at our disposal; and the system that we construct is essentially a presentation that we make to each other, not a discourse by the subject. In the case above, if (C) is true, and even if (A) and (B) are tautologies, we still may wonder whether or not our subject, *a*, knows (A) and (B) and knows the force of the substitutions suggested by them.

In his presentation of epistemic logic, D. Paul Snyder uses the phrase "responsible knowing," for the case in which "one is responsible for the logical consequences of what one knows," and all truths of logic are supposed to be known.⁶ We will adopt a somewhat similar notion of *responsible ignorance* in which one is taken to be responsible for the logical consequences of his ignorance. Where the subject is not ignorant of a particular consequence, that can be separately stated. This notion of responsible ignorance is something like the "alternative interpretation" that Hintikka considers but rejects.⁷ He mentions that one might take the course of reading his "*K_ap*" as "It follows from what *a* knows that *p*." However he insists that his notions of defensibility and virtual implication give something quite different. Whatever the case, in our *G(a, p)* we may attribute to *a* or deny to him whatever logically follows from the ignorance of *p*. What does follow from ignorance of a sentence? A formal answer requires a careful statement of the machinery for introduction and elimination, etc. To spare the detour, we will adopt Quine's *M.L.*⁸ for the examiner's language. Let us define a responsible ignoramus (a subject) as one whose ignorance is consistent with the results of *M.L.* or an equivalent system of rules. First, we need a symbol for logical consequence. Take

D 10 $p \vdash q =_{df} q$ is a logical consequence of p (under *M.L.* or an equivalent system of rules).

Then

Definition 11 *a* is a responsible ignoramus =_{df}

$$[p \& (p \vdash q)] \supset [G(a, q) \supset \neg G(a, p)]$$

For special purposes we might assume an ignoramus with different logical powers and maturity. Our responsible ignoramus is responsible for quite a large degree of logical power; in many applications one might want to weaken the logical ability a great deal. In the case of our Definition 11, we have a subject that is not capable of knowing contradictions, or denying tautologies, etc. And, our subject is responsible for ignorance of any statement that is logically equivalent to an already established matter of ignorance for him.

An individual with the logical nature assumed above cannot *know* that contradictories are both true or both false. Are his possible patterns of ignorance equally restricted? (1) states that he either is or is not ignorant of the reference of sentence *s*. We can add to this

$$T (1.1) \quad G^*(a, s) \vee \neg G^*(a, s)$$

and

$$T (1.2) \quad G^v(a, s) \vee \neg G^v(a, s).$$

But, can he be ignorant of both parts of a contradictory pair? Consider

$$(D) \quad G^*(a, \neg s) \& G^*(a, s)$$

$$(E) \quad G^v(a, s) \& G^v(a, \neg s)$$

$$(F) \quad G(a, s) \& G(a, \neg s)$$

(D) is certainly possible, since a rational ignoramus knows that any statement can be denied: if he were unable to mention '*s*' he would be unable to mention '*¬s*'. From either side of the conjunction we can infer the other, giving

$$T (1.3) \quad G^*(a, s) \equiv G^*(a, \neg s).$$

(E) asserts that *a* is ignorant of the truth-value of '*s*' and of the truth value of '*¬s*'. Again, a rational ignoramus knows that a denial sign in front of a statement reverses its truth-value. So, if he knew the truth-value of either '*s*' or '*¬s*' he would know the other. To be ignorant of either one is to be ignorant of the other, giving

$$T (1.4) \quad G^v(a, s) \equiv G^v(a, \neg s).$$

⁶ D. P. Snyder, *Modal Logic and Its Applications* (New York, 1971), p. 201.

⁷ Jaakko Hintikka, *Knowledge and Belief* (Ithaca, 1962), p. 38-39.

⁸ Cited in n. 2.

(F) seems deviant. Where s is a sentence asserting the fact that Lincoln was assassinated, (F) asserts that a is ignorant of the fact that he was and ignorant of the fact that he was not assassinated. This would be entirely too much ignorance for a rational ignoramus. One and only one side of the conjunction can occur, as we understand the terms involved. Ordinary usage is not perturbed by the fact that even one who was assassinated must have had many unassassinated moments. If someone, say Washington, was not assassinated, there can be abnorance but not ignorance of his assassination. So, differing from (D) and (E), in (F) both sides of the conjunction cannot be true. Here we have another distinction between ignorance of reference and the other two kinds of ignorance. Of course, there can be ignorance of facts; in addition to (I) we can have

T (6) $G^v(a, s) \supset [G(a, s) \vee G(a, -s)]$

with the examiner knowing that one side of the alternation is false.

Axiom III seems to permit an alarming result. This axiom gives the expected

T (7) $G(a, s) \supset s$,

but by transposition, apparently

(G) $-s \supset -G(a, s)$.

(G) looks as if it asserts that if s is false then a knows that it is true. However, the apparent paradox of (G) is simply the result of the error of misapplying the Law of transposition. Brackets around the left side of the connective in (7) must be understood, and will avoid the pitfall.

From (7) we obviously obtain

T (7.1) $-s \supset -[G(a, s)]$.

This holds that if s is false, it is false to hold that a is ignorant of s . Perfectly acceptable, while (G) was not. We must distinguish between

(H) It is false to say that a is ignorant of s ,
and

(I) a is not ignorant of s .

What difference is there between (H) and (I)?

(H) is clear, and (I) is ambiguous, about the position of its negation. We can maintain definitions 7 and 8 for K , and avoid the paradox of (G) provided we are scrupulous in applying the laws of statement logic.

At this state it is natural to wonder what can be obtained for distribution, iteration, quantification and image patterns. A few samples follow. Snyder's T6 for his epistemic operator⁹ holds that when there is knowledge of a conjunction, there is independent knowledge of each part. For the counter-epistemic operator, however, the consequence can only be an alternation, as

T (8) $G^*(a, p \& q) \supset [G^*(a, p) \vee G^*(a, q)]$.

If both p and q could be mentioned, a responsible ignoramus could conjoin them. Also, a responsible ignoramus who is able to mention ' $p \& q$ ' would be able to apply the law of simplification and obtain either part of the conjunction separately.

T (8.1) $-G^*(a, p \& q) \supset [-G^*(a, p) \& -G^*(a, q)]$.

Since our ignoramus is responsible for knowing all tautologies, we expect that

T (9) $G^v(a, p \supset q) \supset [G^v(a, p) \vee G^v(a, q)]$

If a is ignorant of the truth-value of a conditional, either he is ignorant of the truth-value of the antecedent or of the consequent. (9) cannot be strengthened into an equivalence because the subject might be ignorant of just one part of the alternation, know the other, and not be ignorant of the conditional. The possibilities of quantification are enormous but treacherous. Some obvious relations are

T (10) $(\exists x)G(x, s) \supset s$

T (11) $G(a, s) \supset (\exists x)G(x, s)$.

These last two, together, give us (7), of course.

As mentioned earlier, this semantic conception of ignorance is an image relation. The G image of L are those who are ignorant of sentences that are members of the class L ,

T (12) $\hat{x} (\exists y)[G(x, y) \& (y \in L)]$.

After Quine's theorem 469,¹⁰

T (13) $(G^v \subset G^*) \supset [(G^{v''s}) \subset (G^{*''s})]$.

If ignorance of truth-value is included in ignorance of mention, then ignorance of the truth-value of a sentence is included in ignorance of the mention of the sentence. Fully quantified, (13) is

T (13.1) $(\forall y)(\forall x)(\forall z)(x \subset y) \supset [(x''z) \subset (y''z)]$.

After Quine's 470 (*op cit*), If French is a language,

⁹ D. P. Snyder, *op. cit.*, p. 203.

¹⁰ W. V. O. Quine, *Mathematical Logic, op. cit.*, p. 210.

then ignorance of French is ignorance of a language. With obvious abbreviations,

T (14) $(F \subset L) \supset [(G^*F) \subset (G^*L)]$.

Quine offers two distributive theorems that can serve our purposes with some alterations. His 471 and 472 suggest (15) and (16).

T (15) $[(G^* \vee G^v)^*s] \equiv [(G^*s) \vee (G^vs)]$.

Ignorance of the mention or the truth value of a sentence is equivalent to ignorance of the truth value of that sentence or ignorance of the mention of that sentence.

T (16) $[G^*(s^1 \vee s^2)] \equiv [(G^*s^1) \vee (G^*s^2)]$.

Ignorance of either one sentence or a second is equivalent to ignorance of the first or ignorance of the second. (14), (15), and (16) can also be fully quantified in a straightforward way.

What Quine calls the *converse image*¹¹ may also be used for our purposes. It is the class of all statements to which a certain class of individuals bear the ignorance relation. The applications are obvious and will not be developed here.

In epistemic logic the question of forgetting has raised problems. For Hintikka, "only statements made on one and the same occasion are considered here."¹² For Snyder, "forgetting is not allowed."¹³ We suggest two axioms to hold that if something is known it is known for some specific interval thereafter; and if an individual is ignorant of something, he was ignorant of it for some interval preceding a specified time. Let the time elements be represented by real numbers:

A VII. $K(a, s, t) \supset K(a, s, (t + |x|))$.

A VIII. $G(a, s, t) \supset G(a, s, (t - |x|))$.

We suggest that VII and VIII be called the elephant axioms, since elephants are taken to satisfy them for large values of x .

Now we turn to considering certain iterations. There is a serious difference between

(J) $G^v(a, s) \& K(a, G^v(a, s))$,

and

(K) $G^v(a, s) \& G^v(a, G^v(a, s))$.

Take 's' in (J) and (K) to stand for "The sentence, 'Water contains oxygen' is true." Certainly (J) can occur: one can know that, poor chemist that he is, he is ignorant of the truth-value of that sentence

about water and that other thing. It is even easier to illustrate (K).

Can one know that he is ignorant of the mention of a sentence? Visitors to foreign countries are occasionally in the predicament of knowing that there is some sentence on a particular subject, but they do not have the ability to mention such a sentence. The observer who is sufficiently knowledgeable can express the situation more specifically. For the subject concerned, a , it would be true that

(L) $(\exists x)[(x \text{ is a sentence giving direction for reaching the airport}) \& K(a, G^*(a, x))]$.

The observer might also be able to specify what could satisfy x .

Can one know that he is ignorant of a fact? From the examples above, he can know that he is ignorant of the mention of a sentence asserting a fact, and of the truth-value of a sentence asserting a fact. But, to know that he is ignorant of a fact he cannot, of course, assert a description of the fact but must instead assert that there is some fact in a certain area or on a certain topic. Perhaps

(M) $K(a, [(\exists x)(x \text{ is a sentence about the composition of the water molecule}) \& G(a, x)])$.

Of course the possibilities for iterated ignorance are enormous. We have no trouble finding cases of

(N) $G(a, G(a, s))$.

As long as we are careful to distinguish between ignorance of mention, of truth-value, and of reference, there is no difficulty in producing additional theorems involving iteration. Socratic ignorance can take the form of (J), (L), or (M).

To assert that an adequate statement on some matter does not exist (that a problem cannot be solved), one may hold that no one knows the pertinent sentence and no one is ignorant of the sentence.

T (17) $(\forall x) [-K(x, s) \& -G(x, s)] \supset -(\exists y)$
($y = s$)

would mean that there is no such sentence. A better way of expressing it would be by the three-place relation, *abnornance*, defined earlier. Abnornance holds when a is unable to assert s because s does not exist in his language. For example, let s_1

¹¹ *Ibid.*, p. 212.

¹² *Op. cit.*, p. 106.

¹³ *Op. cit.*, p. 202.

be the correct formula for the area of a square-circle. Then,

T (18) $(\forall t) (\forall x) B(x, s_L, t)$

Of course, the observers can be in error, can take something to be abnornance when they are merely projecting their own ignorance.

Distinguishing between kinds of ignorance can help in considering the paradoxes. Let 'u' be a so-called undecidable sentence (e.g., "This sentence is false"). Then, ignorance of the mention of u can occur, but not ignorance of its truth-value. Of course, there can be ignorance of just what happens to different assignments of truth-value.

A number of elementary patterns have some interest, such as

T (19) $[G(a, q) \& [K(a, p) \supset K(a, q)]] \supset G(a, p)$

by *modus tollens*, and

T (20) $[G(a, p) \& [G(a, p) \supset G(a, q)]] \supset G(a, q)$

by *modus ponens*. If, as in (19), a is ignorant of q, and he knows p only if he knows q; then he is ignorant of p. Similarly, in (20), if a is ignorant of p, and if when he is ignorant of p he is ignorant of q; he is ignorant of q.

6. Concluding Comments

To clarify the applications of the notion of semantic ignorance, we offer some behavioral examples. Let j stand for "Jones" and s stand for the sentence, "water contains oxygen." $G(j, s)$ describes the case in which Jones cannot produce the sentence, "water contains oxygen," when asked to mention all the things chemically contained in water, etc. The fair examiner may grant some degree of responsible ignorance and accept close equivalents. If he cannot assert s under what we take to be relevant examination conditions, he is apparently ignorant of the sentence. Also, if asked some quite different question and he asserts s, he then is not ignorant of every aspect of the sentence: he can mention it, although he may or may not know its truth-value or its reference. Likewise, if he can sing a song containing the sentence, he has the ability to mention it, and so is not completely ignorant of it. If our Jones can physically produce, locate, or communicate the sentence, he is not ignorant of its mention. We must distinguish this from the case in which one might produce some

sentence but not know that one has made a statement . . . as a child singing a song or a bit of doggerel in a language foreign to it might take the words to be merely sounds rather than elements or sentences of a language.

Suppose that Jones does have a demonstrated knowledge of the words "water," "contains," and "oxygen," and enough grammar to have made them into a well-formed sentence . . . but never has done so. Can we hold him ignorant of the mention of the sentence? The issue is whether or not he deals with the words as a sentence; if under *fair examination conditions* he fails to produce the words as a sentence, we take him to be ignorant of the mention, truth-value or reference of it. If we design fair examination conditions, we can decide that Jones is or is not ignorant. Can we be charged with restricting ourselves to the examined Jones, rather than Jones-in-himself? Only by those who can point out a way of measuring the ignorance of that last entity.

One might be able to mention a sentence, and know its truth-value, but not know what situation is described by it. By carrying out the steps of a proof in mathematics, or by substitution and rearrangement of symbols in some branch of science, one might believe that a sentence is true but not know what it referred to. The history of science has such examples.

The relationship between knowledge, ignorance, and abnornance is not always obvious. Even if an individual adds to his knowledge, he may also add to one or another kind of ignorance. He may forget, as well. The increase of knowledge may disclose problems of which he was ignorant. He may move from ignorance of the existence of a sentence to ignorance of its truth-value or its reference.

Whether or not ignorance is always bad is a moot point. Attention could be called to one classical argument for the moral requirement of a certain kind of ignorance. Kant argues¹⁴ that morality requires ignorance of the truth-value of "God exists." In our terms, he would not specify ignorance; either that or abnornance, whatever the case may be. In his *Introduction to Logic*,¹⁵ Kant distinguished between four kinds of ignorance: what we cannot know, dare not know, need not know, and ought not know. It would seem obvious that for scientific and other purposes, a precise analysis of ignorance is a matter of significance.

¹⁴ *Critique of Practical Reason* (Prussian Academy Edition), p. 147.

¹⁵ T. K. Abbott (trs.), *Kant's Introduction to Logic* (London, 1885), pp. 32-33.

- "Ignorance of the law," and "Ignorance of right and wrong," are examples of matters that could benefit from analysis using distinctions of the sort that we have offered here.¹⁸

Temple University
The National Law Center, George Washington University

Received September 23, 1974

¹⁸ We acknowledge with thanks the comments by Hugues Leblanc and William Wisdom on an early draft, and by John C. Nelson and an anonymous referee of this journal on the present version.

VI. THE PARADOXES OF TIME TRAVEL

DAVID LEWIS

TIME travel, I maintain, is possible. The paradoxes of time travel are oddities, not impossibilities. They prove only this much, which few would have doubted: that a possible world where time travel took place would be a most strange world, different in fundamental ways from the world we think is ours.

I shall be concerned here with the sort of time travel that is recounted in science fiction. Not all science fiction writers are clear-headed, to be sure, and inconsistent time travel stories have often been written. But some writers have thought the problems through with great care, and their stories are perfectly consistent.¹

If I can defend the consistency of some science fiction stories of time travel, then I suppose parallel defenses might be given of some controversial physical hypotheses, such as the hypothesis that time is circular or the hypothesis that there are particles that travel faster than light. But I shall not explore these parallels here.

What is time travel? Inevitably, it involves a discrepancy between time and time. Any traveler departs and then arrives at his destination; the time elapsed from departure to arrival (positive, or perhaps zero) is the duration of the journey. But if he is a time traveler, the separation in time between departure and arrival does not equal the duration of his journey. He departs; he travels for an hour, let us say; then he arrives. The time he reaches is not the time one hour after his departure. It is later, if he has traveled toward the future; earlier, if he has traveled toward the past. If he has traveled far toward the past, it is earlier even than his departure. How can it be that the same two events, his departure and his arrival, are separated by two unequal amounts of time?

It is tempting to reply that there must be two independent time dimensions; that for time travel to be possible, time must be not a line but a plane.² Then a pair of events may have two unequal separations if they are separated more in one of the time dimensions than in the other. The lives of common people occupy straight diagonal lines across the plane of time, sloping at a rate of exactly one hour of time₁ per hour of time₂. The life of the time traveler occupies a bent path, of varying slope.

On closer inspection, however, this account seems not to give us time travel as we know it from the stories. When the traveler revisits the days of his childhood, will his playmates be there to meet him? No; he has not reached the part of the plane of time where they are. He is no longer separated from them along one of the two dimensions of time, but he is still separated from them along the other. I do not say that two-dimensional time is impossible, or that there is no way to square it with the usual conception of what time travel would be like. Nevertheless I shall say no more about two-dimensional time. Let us set it aside, and see how time travel is possible even in one-dimensional time.

The world—the time traveler's world, or ours—is a four-dimensional manifold of events. Time is one dimension of the four, like the spatial dimensions except that the prevailing laws of nature discriminate between time and the others—or rather, perhaps, between various timelike dimensions and various spacelike dimensions. (Time remains one-dimensional, since no two timelike dimensions are orthogonal.) Enduring things are timelike streaks: wholes composed of temporal parts, or *stages*, located at various times and places. Change is qualitative differ-

¹ I have particularly in mind two of the time travel stories of Robert A. Heinlein: "By His Bootstrap;" in R. A. Heinlein, *The Menace from Earth* (Hicksville, N.Y., 1959), and "—All You Zombies—," in R. A. Heinlein, *The Unpleasant Profession of Jonathan Hoag* (Hicksville, N.Y., 1959).

² Accounts of time travel in two-dimensional time are found in Jack W. Meiland, "A Two-Dimensional Passage Model of Time for Time Travel," *Philosophical Studies*, vol. 26 (1974), pp. 153–173; and in the initial chapters of Isaac Asimov, *The End of Eternity* (Garden City, N.Y., 1955). Asimov's denouement, however, seems to require some different conception of time travel.

ence between different stages—different temporal parts—of some enduring thing, just as a “change” in scenery from east to west is a qualitative difference between the eastern and western spatial parts of the landscape. If this paper should change your mind about the possibility of time travel, there will be a difference of opinion between two different temporal parts of you, the stage that started reading and the subsequent stage that finishes.

If change is qualitative difference between temporal parts of something, then what doesn't have temporal parts can't change. For instance, numbers can't change; nor can the events of any moment of time, since they cannot be subdivided into dissimilar temporal parts. (We have set aside the case of two-dimensional time, and hence the possibility that an event might be momentary along one time dimension but divisible along the other.) It is essential to distinguish change from “Cambridge change,” which can befall anything. Even a number can “change” from being to not being the rate of exchange between pounds and dollars. Even a momentary event can “change” from being a year ago to being a year and a day ago, or from being forgotten to being remembered. But these are not genuine changes. Not just any old reversal in truth value of a time-sensitive sentence about something makes a change in the thing itself.

A time traveler, like anyone else, is a streak through the manifold of space-time, a whole composed of stages located at various times and places. But he is not a streak like other streaks. If he travels toward the past he is a zig-zag streak, doubling back on himself. If he travels toward the future, he is a stretched-out streak. And if he travels either way instantaneously, so that there are no intermediate stages between the stage that departs and the stage that arrives and his journey has zero duration, then he is a broken streak.

I asked how it could be that the same two events were separated by two unequal amounts of time, and I set aside the reply that time might have two independent dimensions. Instead I reply by distinguishing time itself, *external time* as I shall also call it, from the *personal time* of a particular time traveler: roughly, that which is measured by his wristwatch. His journey takes an hour of his personal time, let us say; his wristwatch reads an hour later at arrival than at

departure. But the arrival is more than an hour after the departure in external time, if he travels toward the future; or the arrival is before the departure in external time (or less than an hour after), if he travels toward the past.

That is only rough. I do not wish to define personal time operationally, making wristwatches infallible by definition. That which is measured by my own wristwatch often disagrees with external time, yet I am no time traveler; what my misregulated wristwatch measures is neither time itself nor my personal time. Instead of an operational definition, we need a functional definition of personal time: it is that which occupies a certain role in the pattern of events that comprise the time traveler's life. If you take the stages of a common person, they manifest certain regularities with respect to external time. Properties change continuously as you go along, for the most part, and in familiar ways. First come infantile stages. Last come senile ones. Memories accumulate. Food digests. Hair grows. Wristwatch hands move. If you take the stages of a time traveler instead, they do not manifest the common regularities with respect to external time. But there is one way to assign coordinates to the time traveler's stages, and one way only (apart from the arbitrary choice of a zero point), so that the regularities that hold with respect to this assignment match those that commonly hold with respect to external time. With respect to the correct assignment properties change continuously as you go along, for the most part, and in familiar ways. First come infantile stages. Last come senile ones. Memories accumulate. Food digests. Hair grows. Wristwatch hands move. The assignment of coordinates that yields this match is the time traveler's personal time. It isn't really time, but it plays the role in his life that time plays in the life of a common person. It's enough like time so that we can—with due caution—transplant our temporal vocabulary to it in discussing his affairs. We can say without contradiction, as the time traveler prepares to set out, “Soon he will be in the past.” We mean that a stage of him is slightly later in his personal time, but much earlier in external time, than the stage of him that is present as we say the sentence.

We may assign locations in the time traveler's personal time not only to his stages themselves but also to the events that go on around him.

Soon Caesar will die, long ago; that is, a stage slightly later in the time traveler's personal time than his present stage, but long ago in external time, is simultaneous with Caesar's death. We could even extend the assignment of personal time to events that are not part of the time traveler's life, and not simultaneous with any of his stages. If his funeral in ancient Egypt is separated from his death by three days of external time and his death is separated from his birth by three score years and ten of his personal time, then we may add the two intervals and say that his funeral follows his birth by three score years and ten and three days of *extended personal time*. Likewise a bystander might truly say, three years after the last departure of another famous time traveler, that "he may even now—if I may use the phrase—be wandering on some plesiosaurus-haunted oolitic coral reef, or beside the lonely saline seas of the Triassic Age."³ If the time traveler does wander on an oolitic coral reef three years after his departure in his personal time, then it is no mistake to say with respect to his extended personal time that the wandering is taking place "even now".

We may liken intervals of external time to distances as the crow flies, and intervals of personal time to distances along a winding path. The time traveler's life is like a mountain railway. The place two miles due east of here may also be nine miles down the line, in the west-bound direction. Clearly we are not dealing here with two independent dimensions. Just as distance along the railway is not a fourth spatial dimension, so a time traveler's personal time is not a second dimension of time. How far down the line some place is depends on its location in three-dimensional space, and likewise the locations of events in personal time depend on their locations in one-dimensional external time.

Five miles down the line from here is a place where the line goes under a trestle; two miles further is a place where the line goes over a trestle; these places are one and the same. The trestle by which the line crosses over itself has two different locations along the line, five miles down from here and also seven. In the same way, an event in a time traveler's life may have more than one location in his personal time. If

he doubles back toward the past, but not too far, he may be able to talk to himself. The conversation involves two of his stages, separated in his personal time but simultaneous in external time. The location of the conversation in personal time should be the location of the stage involved in it. But there are two such stages; to share the locations of both, the conversation must be assigned two different locations in personal time.

The more we extend the assignment of personal time outwards from the time traveler's stages to the surrounding events, the more will such events acquire multiple locations. It may happen also, as we have already seen, that events that are not simultaneous in external time will be assigned the same location in personal time—or rather, that at least one of the locations of one will be the same as at least one of the locations of the other. So extension must not be carried too far, lest the location of events in extended personal time lose its utility as a means of keeping track of their roles in the time traveler's history.

A time traveler who talks to himself, on the telephone perhaps, looks for all the world like two different people talking to each other. It isn't quite right to say that the whole of him is in two places at once, since neither of the two stages involved in the conversation is the whole of him, or even the whole of the part of him that is located at the (external) time of the conversation. What's true is that he, unlike the rest of us, has two different complete stages located at the same time at different places. What reason have I, then, to regard him as one person and not two? What unites his stages, including the simultaneous ones, into a single person? The problem of personal identity is especially acute if he is the sort of time traveler whose journeys are instantaneous, a broken streak consisting of several unconnected segments. Then the natural way to regard him as more than one person is to take each segment as a different person. No one of them is a time traveler, and the peculiarity of the situation comes to this: all but one of these several people vanish into thin air, all but another one appear out of thin air, and there are remarkable resemblances between one at his

³ H. G. Wells, *The Time Machine, An Invention* (London, 1895), epilogue. The passage is criticized as contradictory in Donald C. Williams, "The Myth of Passage," *The Journal of Philosophy*, vol. 48 (1951), p. 463.

appearance and another at his vanishing. Why isn't that at least as good a description as the one I gave, on which the several segments are all parts of one time traveler?

I answer that what unites the stages (or segments) of a time traveler is the same sort of mental, or mostly mental, continuity and connectedness that unites anyone else. The only difference is that whereas a common person is connected and continuous with respect to external time, the time traveler is connected and continuous only with respect to his own personal time. Taking the stages in order, mental (and bodily) change is mostly gradual rather than sudden, and at no point is there sudden change in too many different respects all at once. (We can include position in external time among the respects we keep track of, if we like. It may change discontinuously with respect to personal time if not too much else changes discontinuously along with it.) Moreover, there is not too much change altogether. Plenty of traits and traces last a lifetime. Finally, the connectedness and the continuity are not accidental. They are explicable; and further, they are explained by the fact that the properties of each stage depend causally on those of the stages just before in personal time, the dependence being such as tends to keep things the same.⁴

To see the purpose of my final requirement of causal continuity, let us see how it excludes a case of counterfeit time travel. Fred was created out of thin air, as if in the midst of life; he lived a while, then died. He was created by a demon, and the demon had chosen at random what Fred was to be like at the moment of his creation. Much later someone else, Sam, came to resemble Fred as he was when first created. At the very moment when the resemblance became perfect, the demon destroyed Sam. Fred and Sam together are very much like a single person: a time traveler whose personal time starts at Sam's birth, goes on to Sam's destruction and Fred's creation, and goes on from there to Fred's death. Taken in this order, the stages of Fred-cum-Sam have the proper connectedness and continuity. But they lack causal continuity, so Fred-cum-Sam is not one person and not a time traveler.

Perhaps it was pure coincidence that Fred at his creation and Sam at his destruction were exactly alike; then the connectedness and continuity of Fred-cum-Sam across the crucial point are accidental. Perhaps instead the demon remembered what Fred was like, guided Sam toward perfect resemblance, watched his progress, and destroyed him at the right moment. Then the connectedness and continuity of Fred-cum-Sam has a causal explanation, but of the wrong sort. Either way, Fred's first stages do not depend causally for their properties on Sam's last stages. So the case of Fred and Sam is rightly disqualified as a case of personal identity and as a case of time travel.

We might expect that when a time traveler visits the past there will be reversals of causation. You may punch his face before he leaves, causing his eye to blacken centuries ago. Indeed, travel into the past necessarily involves reversed causation. For time travel requires personal identity—he who arrives must be the same person who departed. That requires causal continuity, in which causation runs from earlier to later stages in the order of personal time. But the orders of personal and external time disagree at some point, and there we have causation that runs from later to earlier stages in the order of external time. Elsewhere I have given an analysis of causation in terms of chains of counterfactual dependence, and I took care that my analysis would not rule out causal reversal *a priori*.⁵ I think I can argue (but not here) that under my analysis the direction of counterfactual dependence and causation is governed by the direction of other *de facto* asymmetries of time. If so, then reversed causation and time travel are not excluded altogether, but can occur only where there are local exceptions to these asymmetries. As I said at the outset, the time traveler's world would be a most strange one.

Stranger still, if there are local—but only local—causal reversals, then there may also be causal loops: closed causal chains in which some of the causal links are normal in direction and others are reversed. (Perhaps there must be loops if there is reversal; I am not sure.) Each event on the loop has a causal explanation, being caused by events elsewhere on the loop. That is not to

⁴ I discuss the relation between personal identity and mental connectedness and continuity at greater length in "Survival and Identity" in *The Identity of Persons*, ed. by Amelie Rorty (forthcoming).

⁵ "Causation," *The Journal of Philosophy*, vol. 70 (1973), pp. 556–567; the analysis relies on the analysis of counterfactuals given in my *Counterfactuals* (Oxford, 1973).

say that the loop as a whole is caused or explicable. It may not be. Its inexplicability is especially remarkable if it is made up of the sort of causal processes that transmit information. Recall the time traveler who talked to himself. He talked to himself about time travel, and in the course of the conversation his older self told his younger self how to build a time machine. That information was available in no other way. His older self knew how because his younger self had been told and the information had been preserved by the causal processes that constitute recording, storage, and retrieval of memory traces. His younger self knew, after the conversation, because his older self had known and the information had been preserved by the causal processes that constitute telling. But where did the information come from in the first place? Why did the whole affair happen? There is simply no answer. The parts of the loop are explicable, the whole of it is not. Strange! But not impossible, and not too different from inexplicabilities we are already inured to. Almost everyone agrees that God, or the Big Bang, or the entire infinite past of the universe, or the decay of a tritium atom, is uncaused and inexplicable. Then if these are possible, why not also the inexplicable causal loops that arise in time travel?

I have committed a circularity in order not to talk about too much at once, and this is a good place to set it right. In explaining personal time, I presupposed that we were entitled to regard certain stages as comprising a single person. Then in explaining what united the stages into a single person, I presupposed that we were given a personal time order for them. The proper way to proceed is to define personhood and personal time simultaneously, as follows. Suppose given a pair of an aggregate of person-stages, regarded as a candidate for personhood, and an assignment of coordinates to those stages, regarded as a candidate for his personal time. If the stages satisfy the conditions given in my circular explanation with respect to the assignment of coordinates, then both candidates succeed: the stages do comprise a person and the assignment is his personal time.

I have argued so far that what goes on in a time travel story may be a possible pattern of events in four-dimensional space-time with no extra time dimension; that it may be correct to

regard the scattered stages of the alleged time traveler as comprising a single person; and that we may legitimately assign to those stages and their surroundings a personal time order that disagrees sometimes with their order in external time. Some might concede all this, but protest that the impossibility of time travel is revealed after all when we ask not what the time traveler *does*, but what he *could do*. Could a time traveler change the past? It seems not: the events of a past moment could no more change than numbers could. Yet it seems that he would be as able as anyone to do things that would change the past if he did them. If a time traveler visiting the past both could and couldn't do something that would change it, then there cannot possibly be such a time traveler.

Consider Tim. He detests his grandfather, whose success in the munitions trade built the family fortune that paid for Tim's time machine. Tim would like nothing so much as to kill Grandfather, but alas he is too late. Grandfather died in his bed in 1957, while Tim was a young boy. But when Tim has built his time machine and traveled to 1920, suddenly he realizes that he is not too late after all. He buys a rifle; he spends long hours in target practice; he shadows Grandfather to learn the route of his daily walk to the munitions works; he rents a room along the route; and there he lurks, one winter day in 1921, rifle loaded, hate in his heart, as Grandfather walks closer, closer, . . .

Tim can kill Grandfather. He has what it takes. Conditions are perfect in every way: the best rifle money could buy, Grandfather an easy target only twenty yards away, not a breeze, door securely locked against intruders, Tim a good shot to begin with and now at the peak of training, and so on. What's to stop him? The forces of logic will not stay his hand! No powerful chaperone stands by to defend the past from interference. (To imagine such a chaperone, as some authors do, is a boring evasion, not needed to make Tim's story consistent.) In short, Tim is as much able to kill Grandfather as anyone ever is to kill anyone. Suppose that down the street another sniper, Tom, lurks waiting for another victim, Grandfather's partner Tom is not a time traveler, but otherwise he is just like Tim: same make of rifle, same murderous intent, same everything. We can even suppose that Tom, like Tim, believes himself to be a time traveler.

Someone has gone to a lot of trouble to deceive Tom into thinking so. There's no doubt that Tom can kill his victim; and Tim has everything going for him that Tom does. By any ordinary standards of ability, Tim can kill Grandfather.

Tim cannot kill Grandfather. Grandfather lived, so to kill him would be to change the past. But the events of a past moment are not subdivisible into temporal parts and therefore cannot change. Either the events of 1921 timelessly do include Tim's killing of Grandfather, or else they timelessly don't. We may be tempted to speak of the "original" 1921 that lies in Tim's personal past, many years before his birth, in which Grandfather lived; and of the "new" 1921 in which Tim now finds himself waiting in ambush to kill Grandfather. But if we do speak so, we merely confer two names on one thing. The events of 1921 are doubly located in Tim's (extended) personal time, like the trestle on the railway, but the "original" 1921 and the "new" 1921 are one and the same. If Tim did not kill Grandfather in the "original" 1921, then if he does kill Grandfather in the "new" 1921, he must both kill and not kill Grandfather in 1921—in the one and only 1921, which is both the "new" and the "original" 1921. It is logically impossible that Tim should change the past by killing Grandfather in 1921. So Tim cannot kill Grandfather.

Not that past moments are special; no more can anyone change the present or the future. Present and future momentary events no more have temporal parts than past ones do. You cannot change a present or future event from what it was originally to what it is after you change it. What you *can* do is to change the present or the future from the unactualized way they would have been without some action of yours to the way they actually are. But that is not an actual change: not a difference between two successive actualities. And Tim can certainly do as much; he changes the past from the unactualized way it would have been without him to the one and only way it actually is. To "change" the past in this way, Tim need not do anything momentous; it is enough just to be there, however unobtrusively.

You know, of course, roughly how the story of Tim must go on if it is to be consistent: he somehow fails. Since Tim didn't kill Grandfather in the "original" 1921, consistency demands that

neither does he kill Grandfather in the "new" 1921. Why not? For some commonplace reason. Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy. His failure by no means proves that he was not really able to kill Grandfather. We often try and fail to do what we are able to do. Success at some tasks requires not only ability but also luck, and lack of luck is not a temporary lack of ability. Suppose our other sniper, Tom, fails to kill Grandfather's partner for the same reason, whatever it is, that Tim fails to kill Grandfather. It does not follow that Tom was unable to. No more does it follow in Tim's case that he was unable to do what he did not succeed in doing.

We have this seeming contradiction: "*Tim doesn't, but can, because he has what it takes*" versus "*Tim doesn't, and can't, because it's logically impossible to change the past.*" I reply that there is no contradiction. Both conclusions are true, and for the reasons given. They are compatible because "can" is equivocal.

To say that something can happen means that its happening is compossible with certain facts. Which facts? That is determined, but sometimes not determined well enough, by context. An ape can't speak a human language—say, Finnish—but I can. Facts about the anatomy and operation of the ape's larynx and nervous system are not compossible with his speaking Finnish. The corresponding facts about my larynx and nervous system are compossible with my speaking Finnish. But don't take me along to Helsinki as your interpreter: I can't speak Finnish. My speaking Finnish is compossible with the facts considered so far, but not with further facts about my lack of training. What I can do, relative to one set of facts, I cannot do, relative to another, more inclusive, set. Whenever the context leaves it open which facts are to count as relevant, it is possible to equivocate about whether I can speak Finnish. It is likewise possible to equivocate about whether it is possible for me to speak Finnish, or whether I am able to, or whether I have the ability or capacity or power or potentiality to. Our many words for much the same thing are little help since they do not seem to correspond to different fixed delineations of the relevant facts.

Tim's killing Grandfather that day in 1921 is

compossible with a fairly rich set of facts: the facts about his rifle, his skill and training, the unobstructed line of fire, the locked door and the absence of any chaperone to defend the past, and so on. Indeed it is compossible with all the facts of the sorts we would ordinarily count as relevant in saying what someone can do. It is compossible with all the facts corresponding to those we deem relevant in Tom's case. Relative to these facts, Tim can kill Grandfather. But his killing Grandfather is not compossible with another, more inclusive set of facts. There is the simple fact that Grandfather was not killed. Also there are various other facts about Grandfather's doings after 1921 and their effects: Grandfather begat Father in 1922 and Father begat Tim in 1949. Relative to these facts, Tim cannot kill Grandfather. He can and he can't, but under different delineations of the relevant facts. You can reasonably choose the narrower delineation, and say that he can; or the wider delineation, and say that he can't. But choose. What you mustn't do is waver, say in the same breath that he both can and can't, and then claim that this contradiction proves that time travel is impossible.

Exactly the same goes for Tom's parallel failure. For Tom to kill Grandfather's partner also is compossible with all facts of the sorts we ordinarily count as relevant, but not compossible with a larger set including, for instance, the fact that the intended victim lived until 1934. In Tom's case we are not puzzled. We say without hesitation that he can do it, because we see at once that the facts that are not compossible with his success are facts about the future of the time in question and therefore not the sort of facts we count as relevant in saying what Tom can do.

In Tim's case it is harder to keep track of which facts are relevant. We are accustomed to exclude facts about the future of the time in question, but to include some facts about its past. Our standards do not apply unequivocally to the crucial facts in this special case: Tim's failure, Grandfather's survival, and his subsequent doings. If we have foremost in mind that they lie in the external future of that moment in 1921 when Tim is almost ready to shoot, then we exclude them just as we exclude the parallel facts in Tom's case. But if we have foremost in mind that they precede that moment in Tim's extended personal time, then we tend to include

them. To make the latter be foremost in your mind, I chose to tell Tim's story in the order of his personal time, rather than in the order of external time. The fact of Grandfather's survival until 1957 had already been told before I got to the part of the story about Tim lurking in ambush to kill him in 1921. We must decide, if we can, whether to treat these personally past and externally future facts as if they were straightforwardly past or as if they were straightforwardly future.

Fatalists—the best of them—are philosophers who take facts we count as irrelevant in saying what someone can do, disguise them somehow as facts of a different sort that we count as relevant, and thereby argue that we can do less than we think—indeed, that there is nothing at all that we don't do but can. I am not going to vote Republican next fall. The fatalist argues that, strange to say, I not only won't but can't; for my voting Republican is not compossible with the fact that it was true already in the year 1548 that I was not going to vote Republican 428 years later. My rejoinder is that this is a fact, sure enough; however, it is an irrelevant fact about the future masquerading as a relevant fact about the past, and so should be left out of account in saying what, in any ordinary sense, I can do. We are unlikely to be fooled by the fatalist's methods of disguise in this case or other ordinary cases. But in cases of time travel, precognition, or the like, we're on less familiar ground, so it may take less of a disguise to fool us. Also, new methods of disguise are available, thanks to the device of personal time.

Here's another bit of fatalist trickery. Tim, as he lurks, already knows that he will fail. At least he has the wherewithal to know it if he thinks, he knows it implicitly. For he remembers that Grandfather was alive when he was a boy, he knows that those who are killed are thereafter not alive, he knows (let us suppose) that he is a time traveler who has reached the same 1921 that lies in his personal past, and he ought to understand—as we do—why a time traveler cannot change the past. What is known cannot be false. So his success is not only not compossible with facts that belong to the external future and his personal past, but also is not compossible with the present fact of his knowledge that he will fail. I reply that the fact of his foreknowledge, at the moment while he waits to shoot, is not a

fact entirely about that moment. It may be divided into two parts. There is the fact that he then believes (perhaps only implicitly) that he will fail; and there is the further fact that his belief is correct, and correct not at all by accident, and hence qualifies as an item of knowledge. It is only the latter fact that is not compossible with his success, but it is only the former that is entirely about the moment in question. In calling Tim's state at that moment knowledge, not just belief, facts about personally earlier but externally later moments were smuggled into consideration.

I have argued that Tim's case and Tom's are alike, except that in Tim's case we are more tempted than usual—and with reason—to opt for a semi-fatalist mode of speech. But perhaps they differ in another way. In Tom's case, we can expect a perfectly consistent answer to the counterfactual question: what if Tom had killed Grandfather's partner? Tim's case is more difficult. If Tim had killed Grandfather, it seems offhand that contradictions would have been true. The killing both would and wouldn't have occurred. No Grandfather, no Father; no Father, no Tim; no Tim, no killing. And for good measure: no Grandfather, no family fortune; no fortune, no time machine; no time machine, no killing. So the supposition that Tim killed Grandfather seems impossible in more than the semi-fatalistic sense already granted.

If you suppose Tim to kill Grandfather and hold all the rest of his story fixed, of course you get a contradiction. But likewise if you suppose Tom to kill Grandfather's partner and hold the rest of his story fixed—including the part that told of his failure—you get a contradiction. If you make *any* counterfactual supposition and hold all else fixed you get a contradiction. The thing to do is rather to make the counterfactual supposition and hold all else as close to fixed as you consistently can. That procedure will yield perfectly consistent answers to the question: what if Tim had not killed Grandfather? In that case, some of the story I told would not have

been true. Perhaps Tim might have been the time-traveling grandson of someone else. Perhaps he might have been the grandson of a man killed in 1921 and miraculously resurrected. Perhaps he might have been not a time traveler at all, but rather someone created out of nothing in 1920 equipped with false memories of a personal past that never was. It is hard to say what is the least revision of Tim's story to make it true that Tim kills Grandfather, but certainly the contradictory story in which the killing both does and doesn't occur is not the least revision. Hence it is false (according to the unrevised story) that if Tim had killed Grandfather then contradictions would have been true.

What difference would it make if Tim travels in branching time? Suppose that at the possible world of Tim's story the space-time manifold branches; the branches are separated not in time, and not in space, but in some other way. Tim travels not only in time but also from one branch to another. In one branch Tim is absent from the events of 1921; Grandfather lives; Tim is born, grows up, and vanishes in his time machine. The other branch diverges from the first when Tim turns up in 1920; there Tim kills Grandfather and Grandfather leaves no descendants and no fortune; the events of the two branches differ more and more from that time on. Certainly this is a consistent story; it is a story in which Grandfather both is and isn't killed in 1921 (in the different branches); and it is a story in which Tim, by killing Grandfather, succeeds in preventing his own birth (in one of the branches). But it is not a story in which Tim's killing of Grandfather both does occur and doesn't: it simply does, though it is located in one branch and not the other. And it is not a story in which Tim changes the past. 1921 and later years contain the events of both branches, coexisting somehow without interaction. It remains true at all the personal times of Tim's life, even after the killing, that Grandfather lives in one branch and dies in the other.⁶

Princeton University

Received September 4, 1975

⁶ The present paper summarizes a series of lectures of the same title, given as the Gavin David Young Lectures in Philosophy at the University of Adelaide in July, 1971. I thank the Australian-American Educational Foundation and the American Council of Learned Societies for research support. I am grateful to many friends for comments on earlier versions of this paper; especially Philip Kitcher, William Newton-Smith, J. J. C. Smart, and Donald Williams.

VII. FICTION

J. O. URMSON

THE word "fiction" appears to be not infrequently used in such a way as at least to imply falsehood. Thus, when something is said to be a legal fiction, it is sometimes meant that while it is held to be true in courts of law, it is in fact false. Or, again, when Bishop Butler wrote in the Advertisement to his *The Analogy of Religion* that "it has come, I know not how, to be taken for granted, by many persons, that Christianity is not so much a subject of inquiry; but that it is, now at length, discovered to be fictitious," the allegation he mentions is surely that it is all a fraudulent lie. But it will be the first contention of this paper that what in a work of fiction is properly to be regarded as true or false is for the most part true, while what is fictitious in the primary, and etymological sense of "made up" is properly to be regarded as neither true nor false.

There is, I claim, much that is true either stated or implied in works of fiction. In saying this I am not referring to the alleged revelation of truths, more or less profound, in great works of fiction. I do not wish to discuss this claim, but shall make one comment on it. I think that the skill of the writer of fiction is, as Plato said, not that of the scientist or philosopher. No doubt, to take but one example, Stendhal in *La Chartreuse de Parme* makes clear what it is like to be on the fringe of a great battle better than any military historian; but this is because he is a skilled writer and observer who had had much personal experience of being on the fringe of great battles rather than because he was a master of fiction. But the kind of truth that I have in mind is claimed by me to be present in all fiction, great and trivial, and it is not deep revelation but ordinary everyday truth, for the most part wellknown to all or easily ascertained from any reference book. So whether or not some works of fiction have contributed significantly to our knowledge and understanding of reality, let us leave the question aside.

One sort of truth, or occasionally untruth, which commonly occurs in works of fiction is the

direct statement of the author, *propria persona*, to his reader. "It is not a bad thing to be son-in-law to a potent earl, member of parliament for a county, and possessor of a fine old English seat and a fine old English fortune" Trollope tells us, no doubt with conviction, on an early page of *Doctor Thorne*. The veracity and conviction of the satirically-minded Miss Austen may be less obvious when she begins *Pride and Prejudice* with the announcement that "It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife." But, with whatever malicious intent, it is a direct statement by Miss Austen to her readers.

Nor are these direct communications of the author to the reader confined to sociological generalization. Trollope, and not only he, makes frequent comments on the progress of his story. "It was declared in the early pages of the work that Dr. Thorne was to be our hero; but it would appear very much as though he had latterly been forgotten," he tells us at the beginning of the twenty-third chapter of *Doctor Thorne*. Towards the end of the same book he comments: "But it will be too much, perhaps, to tell the reader what she [Mary Thorne] wore as Beatrice's bridesmaid, seeing that a couple of pages, at least, must be devoted to her own marriage-dress."

It would be a mistake to think that such direct communication by the author to the reader is uncommon to fiction or that when it occurs it can be readily exercised without damage to the main fabric of the work. But no doubt such communication is but one technical device which is employed, little or not at all, in some works of fiction, and we should turn our attention to more essential elements.

Every story must be narrated against a background of presuppositions, since not everything relevant can be stated, and these presuppositions will normally be truths. Thus, if the story begins with the words "Tom was a middle-aged man from Columbus, Ohio," we may assume that he

is visible to other men, needs food and drink, speaks English, and so on, unless we are explicitly warned to the contrary. Without some such presuppositions as these the story will be unintelligible. Even in the most fanciful fairy-stories and science-fiction we, in general, presuppose things to be as they really are except where it is clearly shown or stated to be otherwise. Thus the presupposed is in general the truth, though in certain genres the truth is partially replaced or supplemented by conventional falsehoods; in Wodehouse's novels, for example, it is presupposed that all rich young aristocrats are moronic unless we are explicitly warned of an exception.

Often a single sentence or phrase can give us a very rich background, how rich depending on our state of information. To begin a story with the words "In the winter of 1714 at Fort William..." will allow most potential readers to know that they must presuppose early eighteenth century conditions in a primitive highland area and some to know that they must presuppose an English garrison in a clannish area with the 1715 Jacobite rebellion just round the corner and the Glencoe massacre not far away. But not every reader will know all that he needs to know and it is very likely that some of this will be explicitly stated for the benefit of those who do not. In general what is stated will need to be true if it is not to conflict needlessly with the presuppositions of the better informed.

Absolute adherence to the truth is not always necessary or possible. We have seen already that the exigencies of the story may require a background which departs from it at certain points. Practical considerations such as the law of libel have also to be taken into account as is well illustrated by the foreword to Ngaio Marsh's detective story entitled *Vintage Murder*: "Although I agree with those critics who condemn the building of imaginary towns in actual countries I must confess that there is no Middleton in the North Island of New Zealand, nor is 'Middleton' a pseudonym for any actual city. The largest town in New Zealand is no bigger than, let us say, Southampton. If I had taken the Davis Comedy Company to Auckland or Wellington, Messrs. Wade, Packer and Cass, to say nothing of Dr. Rangi Te Pokiha, might have been mistaken for portraits or caricatures of actual persons."

The setting may be fictional, as Barchester is

a fictional town in the fictional county of Barsetshire. But Trollope wisely tells us that Barsetshire is a typical west-country county of England and Barchester a typical cathedral town, and that the story is set in the mid-nineteenth century—a piece of fiction which enables us to presuppose a rich social background. Even most of the hobbits in *The Lord of the Rings*, if not the heroic Frodo, soon emerge as not unlike Barsetshire characters; while Donald Duck would often be incomprehensible to anyone who could not see him as an American adolescent, typical save for his low intelligence and hot temper.

So, I say, a great deal of truth is inevitably presupposed and implied by all fiction, and this truth is often, though not inevitably, supplemented by social, historical and geographical statements, mostly true.

Closely akin to the presupposed and stated truth, which we need to give the background of the story, is local colour. An enormous amount of true information can be gleaned from novels. One can, for example, learn a lot about the history and geography of Scotland from Scott's novels; all that I know about the manufacture of Scotch whisky I derived from a clearly authoritative and quite detailed account in a quite ordinary romantic novel.

By background truth I mean such truths as one needs to make the story comprehensible; by local colour I mean such matter as lends the story verisimilitude but which is not necessary to its comprehension. I do not claim that the line between them is clear. I also do not wish to claim that the line between them together and the story is always clear. In particular, there is the problem of the historical novel which I must postpone. But by and large we can say "The setting is... and the story is..." without too great difficulty.

With regard to the strategy of this paper we now stand as follows. Having noted that the fictional is sometimes equated with the false, I have shown that, in the form of author's asides, and statements supplying background and local colour, it is certain that in typical works of fiction there are assertions that are true or false. But I have pointed out that of these the majority are true rather than false, and with very good reason. But while these true and false statements may well be integral to the works of fiction in which they occur, it is obvious that it is not in virtue of containing these statements, true or

false, that we would classify the works in which they occur as fiction. I must therefore move on to substantiate by announced contention that the essentially fictional element in works of fiction is neither true nor false.

When Hume said that poets were liars by profession I suppose he was taking for granted the falsity of the fictional. But we shall find ourselves in great difficulties if we allow ourselves to agree with him. Let us suppose that I tell a story to a child and begin in one time-honored style; "There was," I say, "many years ago a little girl who lived with her grandmother on the edge of a large forest." One can scarcely deny that this, if anything, is fiction: but is it false? Perhaps we do not know for certain, so let us ask ourselves whether it is likely that many years ago a little girl lived with her grandmother on the edge of a large forest. If we do ask ourselves this question, the only answer can be that there is an overwhelming statistical probability that it is true. If, in my story, I had said not just "a little girl" but "a little girl named Mary," the statistical probability would still be high, whereas, if I had said "a little girl who was changed by a witch into a water-melon," the statistical probability would no doubt drop to zero. But it is surely too clear to need argument that in raising the question of truth, and resorting to statistical probabilities as the only way of answering the question, we have fallen into utter absurdity. If you ask me whether it is true that a little girl lived many years ago with her grandmother on the edge of a forest, I shall say "true" rather than "false"; but in the imagined context the question should not be raised.

One can tell a child a story and, after finishing the story, say: "and that was a true story—the little girl was your Aunt Tabitha." I could also tell a child a false story, a lie, perhaps by saying: "When your Aunt Tabitha was a little girl she . . ." and then describing a lot of things that never happened to Aunt Tabitha. If, indeed, I tell a story as fiction you may raise the question whether the events narrated could have happened, to which the more plausible answer may be "Yes" or "No." But this is not to say "Probably true" or "Probably false." If a story is told and it is asked whether it is true there are three possible answers: "It is true," "It is false" and "It is fiction," and the third answer is a way of saying that the question does not arise. I assert nothing when I make up a story as fiction, so

a fortiori I do not assert something that is true or false, even by coincidence.

While a story of the form "There was once upon a time a person who did and suffered F_1-F " could perversely be interpreted as a piece of speculative history that might be true, the same could not be with regard to a story beginning as does Jane Austen's *Persuasion*: "Sir Walter Elliott, of Kellynch-hall in Somersetshire, was a man who, for his own amusement, never took up any book but the Baronetage." This is not equivalent to: "There was once a man who was named Sir Walter Elliot. . .". If we could ask Miss Austen to whom she was referring when she wrote the name "Sir Walter Elliot," she might answer "A purely fictional character" or "Nobody." Such answers are quite appropriate, save that they are liable to mislead philosophers, so that, in a philosophical gathering we might prefer her to answer, equally truly, "I was not referring to anybody." Just as "There was once a little girl . . ." has the form of an assertion but is not an assertion, so the opening sentence of *Persuasion* begins with what has the form of a reference but is not a reference. It is not that there is a reference to a fictional object, nor is there the use of a referring expression which fails to secure reference (as when one says "That man over there is tall" when there is no man over there). Jane Austen writes a sentence which has the form of an assertion beginning with a reference, but is in fact neither asserting nor referring; therefore she is not referring to any character, fictional or otherwise, nor does she fail to secure reference, except in the jejune sense in which if I sneeze or open a door I fail to secure reference. Nothing would have counted on this occasion as securing reference, and to suppose it could is to be under the impression that Miss Austen was writing history. Of course one can refer to fictional characters, as one does if one says that Sir Walter Elliot is mentioned at the beginning of *Persuasion*. I do not say that one cannot refer to a fictional character, but that Miss Austen did not on the occasion under discussion.

What I am saying is that making up fiction is not a case of stating, or asserting, or propounding a proposition and includes no acts such as referring. Fiction has the external form of history and biography, but is neither. It is easy enough to find analogies. Thus two men may go through the motions of playing a game of chess, possibly

with complete fidelity to the rules, and not be playing a game of chess. If so, the question who won will not be in a full-blooded sense applicable, even if one ends up with his king in such a position as to be technically check-mated; one cannot win or lose if one is not competing. Our men might go through these moves for a variety of reasons. They may be playing over a game between experts, in which case they will be engaged in some sort of historical investigation. Equally they might do it to interest spectators—they arrange to move in ways which they hope will be interesting, which would have made an exciting contest if it had been a contest; or they might do it to illustrate how a certain opening typically develops. A second analogy will be closer to the case of fiction. I may, in the evening, run over in my mind the events of the day; if this is what I do, then my review is to be judged as true or false. On the other hand I may just day-dream; and now it is inappropriate to ask whether my performance is true or false. A day-dream is neither a report nor a forecast and so is not subject to the criteria to which reports and forecasts are subject. Not to be so subject is what makes it a day-dream. It is as possible to go through the motions of writing history or biography without actually doing so as it is possible to go through the motions of playing chess without actually doing so. In the case of fiction "Is it true?" will be inappropriate for the same reason as "Who won?" is inappropriate to the mock-chess.

If we go through the motions of playing chess we do not make chess moves but move chess-pieces in a way similar to that in which chess moves are made. The mock-chess is also parasitic on real chess in that its interest (if any) will arise from its general resemblance to real chess. (With this point one could compare Wittgenstein's claim that one could not play trains unless one were acquainted with real railroads; though one could, of course, do something physically indistinguishable from playing trains by chance or imitation.) So it is natural to describe the mimic-chess in the language of real chess and speak of a move of the pawn on the queen's file rather than in some more sterilized way. In the same way, though I claim philosophically that *Persuasion* begins with no assertion about anything, one can appropriately say that the opening sentence is about Sir Walter Elliot. No more puzzling is the fact that if I, who have read but did

not write *Persuasion*, say that Sir Walter Elliot read nothing except the *Baronetage*, it will be a true statement verifiable by reading *Persuasion*, whereas the first sentence of *Persuasion*, which runs: "Sir Walter Elliott . . . never took up any book but the *Baronetage*," says nothing true or false. The sentences of fiction have to be distinguished from statements about fiction.

We could here draw a further parallel with portrait-painting. A portrait may be of a person and be held to be a good or bad likeness; this portrait will be analogous to a biography. But a portrait may be of nobody in particular, so that it will be meaningless to ask whether it is a good likeness; in which case it is a mock-portrait, portraying, in a full-blooded sense, nobody. But if the painter cares to label it "Marjory" or "The Pensive Maiden," then the portrait is of Marjory or the pensive maiden in the same way as *Persuasion* is about, *inter alios*, Sir W. Elliott. We could call such a portrait a fictional portrait. This case must be contrasted with a conjectural representation of a yeti or abominable snowman, which, if there be no such beast, will be analogous to a case of attempted but unsuccessful reference.

So we may say that when we speak of a fictitious narrative the adjective "fictitious" has the same general logical character as "toy" has when we speak of a toy locomotive, rather than the logical character of "diesel" in the expression "diesel locomotive"; it indicates that it is not really a narrative at all. Fiction is something which preserves the form of narration, assertion, biography or history, but will be misunderstood if it is taken to be such. That is my thesis, though I have indicated that such quasi-narrative is bound to presuppose much truth and may be mixed with a number of author's asides, background information and local colour which for the most part can be taken to be true and is intended to be.

The main difficulty that I face seems to me to be that of justifying the distinction between genuine description of the background and the quasi-narrative of the story. For most of the time the difficulty is slight. If, in a story set in England, we read: "The Birmingham ring-road system is complicated; no wonder that George and Mary lost their way," the reader will recognize that he is being reminded or informed of a truth about Birmingham against which a fictitious misadventure is placed. But the case is

difficult when we consider such an historical novel as *War and Peace*. In this book such real characters as Napoleon, Alexander and Koutouzov play as important a part as most of the fictional characters and it is hard to distinguish truth from fiction. That Napoleon invaded Russia I know to be true; I know that Napoleon's conversations with fictitious characters must be fictitious. But did Napoleon send the letter to Murat which is given in full in Chapter XLIII, and did he really give the Legion of Honour to a Russian soldier named Lazzerov, as he does in the book? I do not know in either case. Undoubtedly historical events are part of the story, and what is history and what fiction will be in varying degrees unknown to most (perhaps all) readers.

Our problem is not that of the historical novel in the minimum sense of a novel whose setting is a real historical past but of the novel where what is historically true is narrated as part of the story. The embarrassment is that our historical ignorance seems to make it impossible at times to know whether we should treat what we read as genuine narration or as mock-narration.

Part of the answer seems to be that such a book as *War and Peace* is not merely a novel; it is intended as a picture of Russian society in the early nineteenth century, as a study of Napoleon and as a study of the nature of warfare—and no doubt as many other things. In so far as we are concerned with the nature of fiction we may justifiably say that our account does not need to deal adequately with such a book considered under its other aspects. In so far as it is a study of Napoleon its acceptability may well depend to some extent on whether the letter was sent to Murat and whether he gave the Legion of Honour to Lazzerov; looking at the book in that light we should perhaps attempt to find out by historical research. But that, we may say, is not our present concern.

But if we are considering *War and Peace* as a novel does it matter to us whether the Legion of Honour was given on the occasion mentioned, whether it is an element of history introduced or the product of Tolstoy's imagination? I am inclined to answer that where we do not know whether elements in the story are fictitious or based on fact, and it is clearly no part of Tolstoy's intention that we should be able to tell

by reading his book, then reading it as fiction we should treat it as fiction. Such elements should not be treated as assertions or history by those who are reading the book as a novel, even if Tolstoy would be in fact willing to assert their truth and even if their truth or falsehood is relevant to some readings of that ambivalent book. So I think we may say that the historical elements in a novel should be read as mock narrative, and the fact that they have a basis in reality, that they could be asserted as true, is irrelevant.

If the above is, at least in general, correct, then fiction is clearly not as such an art-form but a logically distinct type of utterance. A suitably expressed move in a philosophical discussion may be fiction, as when one says to a student "You are a would-be murderer who has asked me where his intended victim is," proposing to raise the question how Kant would require one to act in such a situation.

No doubt, for a book to be classified as fiction in a library more constraints will be placed on it than that of being of this logical type and of book-length. A logically fictitious account of the formation of a river-delta intended to illustrate the principles of geology and hydrodynamics would not be classified by a librarian as fiction. But fiction in the restricted librarian's sense is also not an art-form as such; it may legitimately have the function of entertaining, of edifying, of embodying a philosophy of history or many other things. It is a vulgar error to suppose that such books are abusing an art-form. But some works of art are fiction. In discussing the nature of fiction I have, therefore, been discussing the nature of one of the media that art may employ.

So, in a way, this paper has been an essay on aesthetics. But, if it is correct, there is one more general lesson to be learnt from it. There are those who assert that sentences are the bearers of truth and falsity, whereas in this paper I have been maintaining that the writer of fiction as such asserts nothing true or false though he employs sentences indistinguishable from those used by those who do. I cannot see how a sentence, however reformulated in logician's language, could be essentially fictional or non-fictional. So those who regard sentences as such as true or false will have to give a very different account from mine of the nature of fiction.

VIII. THE INCREDULOUS HUME

PETER HEATH

AN early nineteenth-century guidebook to Edinburgh preserves, I regret to say, this frivolous and forgotten epitaph to one of that city's greatest men:

Within this monument unblest,
Called vulgarly a tomb,
The impressions and ideas rest
That constituted Hume.

If, contrary to all expectation, the subject of this effusion is currently inhabiting a future state, I hope he will forgive me for quoting it. I also hope that word will have reached him of his impending bicentennial, and that the occasion is pleasing to him. The interval may be an arbitrary one, and from the standpoint of eternity is doubtless quite without meaning. Nobody, I think, ever thought of celebrating such a thing until Garrick did it for Shakespeare in 1769, and, in spite of much ridicule, established a custom that has lasted ever since. Hume, very probably, was one of the scoffers then; but as a lover of literary fame the arrival of his own turn can hardly be altogether unwelcome. A custom, though foolish, can still serve a useful purpose, and a tribute essentially ridiculous may nevertheless be sincere.

Not that Hume is in any need of resurrection. Of all great philosophers he is arguably the most alive, with a reputation much higher, in fact, than it was a century ago. When so many schools still acknowledge his influence, and so many students are still familiar with his writings, the difficulty, rather, is to know what to point to, as the most central and enduring features of his work. A further and notorious difficulty, too, is that the book which now stands highest in public esteem is precisely the one which Hume had the lowest opinion of, and would least have wished to be remembered by. His own candidate for immortality was not the *Treatise*, but the *Enquiry concerning the Principles of Morals*, which posterity has perversely agreed to consider the shallower and less rewarding work. Nor does the choice end there. Two hundred years ago, as

Hume neared his end, the one piece of writing whose preservation and publication most anxiously preoccupied him was the *Dialogues on Natural Religion*; and this, one may suppose, not only because of its literary excellence, but also because it shared with the *Enquiry* a common theme. Though it has not always been fashionable to say so, the evidence suggests that what really mattered to Hume, philosophically, was not his role as a sceptic in epistemology or a naturalist in morals, but rather his position as an enemy to superstition and zealotry, and as an exposé of bad arguments in rational theology. It was these errors, after all, that he considered to be dangerous, whereas those of the philosophers he held to be merely ridiculous. The strength of his feeling on this subject is evinced by many remarks in his letters as well as in his published writings. In turning to consider it we advert, I think, not to a matter of peripheral importance, but to a major source of his influence and a major concern of his life.

The nature of this influence is not seriously in doubt. Hume put paid to the traditional arguments for God's existence, and although much in his critique may not be altogether original, and could scarcely have prevailed, historically, without the later assistance of Darwin and Kant, the fact remains that these arguments are now philosophical museum pieces, and that it was Hume's achievement to have put them on the shelf. Like other museum pieces, they continue, of course, to be interesting objects of study, and occasional attempts are made to refurbish or repair them; but no prudent person would nowadays think of using them; they are much too fragile for that. To many, indeed, the damage done was not at first apparent, owing to the circumspection with which Hume had presented his case. The argument from design, at least, continued to find defenders until well into the following century. The more intelligent of these, however—such as Paley and the authors of the *Bridgewater Treatises*—were aware by then of its inconclusive character, and relied on it to sup-

port, rather than to ground, the edifice of faith. This was more sensible, no doubt, than the attitude of those who mistakenly took Hume to be disputing either the premisses or the conclusion of the argument, and hastened to defend it by noisy iteration of both. But it still suggests a failure to appreciate precisely how dangerous and uncertain in its implications the argument had been shown to be, or the consequent risks of employing it as an adjunct, even, to theistic—let alone Christian—apologetics.

Hume's perfunctory treatment of the *a priori* arguments in the *Dialogues* is attributable, not only to the poor opinion he had of them, but also to his awareness of having sufficiently disposed of them in the *Treatise* and elsewhere. The ontological argument founders on the consideration that existence is not a predicate, not a separate idea, while the analysis of causation gets rid of the cosmological argument by eliminating the sort of real necessity of connection that the inference to a First Mover requires. For good measure, it is pointed out in the *Dialogues* that since the nonexistence of anything is always conceivable, it is impossible to demonstrate of a being that it necessarily exists; for to do so would equally necessarily imply what is false, namely, that its nonexistence entailed a contradiction. The existence of "an eternal succession of objects" again has no need of an extraneous author to ground or explain it. A few terse moves of this kind are enough to fell the *a priori* arguments, leaving the whole burden of the case for theism to be borne by Cleanthes, Hume's identikit exponent of Newtonian natural theology and the *a posteriori* evidences of design.

Hume's strategy against this body of arguments is essentially much simpler than his intricate tactical maneuverings would suggest. In principle he hardly does more than to elaborate ingeniously on the age-old truism that you cannot get something for nothing, or extract a quart from a pint pot. The argument professes to infer an infinitely intelligent, powerful and benign author of nature from the finite appearances of order and teleological contrivance in the observable universe. Such arrangements, in little, invariably bespeak a human contriver; by how much the more should they convince us, in the large, of the hidden presence and activity of a superhuman one. Hume, in dissecting and destroying this persuasive analogy, makes no attempt to deny the supposed facts; he simply argues that

they are consistent with other explanations and other analogies of a less ambitious kind. There is no right to attribute to the causes of such phenomena abilities more extensive than are needed to produce the observed effects; the workings of reason and intelligence, in our ordinary experience, are invariably dependent on a basis of biological organization, not the other way round; the wastage, disorder and suffering in nature are no less relevant as evidence for the impotence, incompetence or malignity of a supposed designer than the beauties and benevolences are for his wisdom, power and goodness; nor will it do to explain such discrepancies away by ascribing them to providential plans that we do not know anything about. In those cases where we argue from a humanly produced effect to the nature and intentions of its unobserved cause, we are guided by the knowledge we already have from past experience of the way such causes work; but, never having been witness to the creation of worlds, we can know or infer nothing about how enterprises on *that* scale are managed or inaugurated. A prudent use of the analogy can get us no further, therefore, than the vague conception of a limited and impersonal demiurge. Such a force or being, even if it existed, would be anything but presentable as an object of worship, and could by no stretch be safely identified with the God of traditional theism.

By these and other arguments, Hume shows conclusively that the analogy with human agency that is relied on in advancing from design in nature to a designer of nature is far too imperfect to lead from a necessarily unique, and perhaps infinite, universe to its equally unique and necessarily infinite architect. The argument is not in any way self-contradictory or logically absurd, but simply too feeble to deliver the desired conclusion: "The consistence is not absolutely denied, only the inference." So far from demonstrating the truth of Christianity, or any other brand of theism, the appeal to design fails even to warrant the most attenuated deism. Its failure demonstrates the futility of attempting to connect the facts of experience with the mysteries of faith or revelation. And from this it follows that such mysteries (if they exist at all) are inevitably impenetrable to reason, which is therefore improperly employed in seeking to probe or consolidate the foundations of religion.

These drastic conclusions, and the arguments supporting them, have been often enough chal-

lenged or disregarded, and as often reiterated and revived. Hume's case, on the whole, still stands. The few small holes that have since been picked in it have not been enough to refute him, and it hardly needs adding that his position has lost little, if anything, by the passage of time and the progress of scientific research. It remains, therefore, if not a permanent accession to human knowledge, at any rate a seemingly irremovable obstacle to certain forms of alluring but unprofitable speculation. In the present state of our knowledge (as Hume is careful to emphasize), we are not equipped to embark on such inquiries by what professes to be a use of the scientific method. That many do so is not denied. It is a fact of human nature that people will be apt to seek an answer to their fears and troubles in explanations of this kind. The name of the process is superstition; its agents are priests; and the influence of the systems of irrational belief so engendered has not, in Hume's opinion, been historically beneficial to mankind. Nevertheless, it is a perfectly natural process, and represents, for the majority, the best they can do to convince themselves, against the run of the evidence, that the world is not altogether hopeless, meaningless or unjust. For the thinking minority, however, the factual claims of religion are seen to be unreasonable, contrary to experience, and in all probability delusive. True religion, for Hume, is thus a bleak and uncomfortable agnosticism, without practical influence or consequences. It is religious only inasmuch as it represents an attitude to life, and one which does not absolutely rule out the "religious hypothesis," however little inclined in its favor. The grounds of its opposition to orthodox dogma are also, in fact, grounds for the rejection of contrary dogmas as well; unlike most of his fellow-infidels, Hume is not, therefore, an atheist or mechanistic materialist. His view is simply that our ignorance affords no presumption of any kind about how the universe originated, or why it should be there.

Hume's critics have been prone to argue that he has no business to write off religious belief as irrational, since all belief, on his own showing, is a product of causes, and not, therefore, a rational state of mind. If all belief is in that sense irrationally conditioned, religion is in the same boat with science, and has at least as good a claim on our acceptance. This is to overlook, however, that Hume is quite well able, on his own principles, to distinguish between reason-

able and unreasonable belief. The beliefs of science are reasonable, insofar as they are first suggested and later confirmed by experience, and have meanwhile been subjected to the sort of systematic testing and sifting which constitutes the scientific method. Other beliefs may be equally natural, as human responses to experience, but if they are not exposed to such tests, or do not survive them, and in general run counter to the other teachings of experience, then it is not reasonable to retain them. Hume's doubts about the rationality of induction and analogy are essentially philosophical doubts; they do not preclude the use, *faute de mieux*, of the ordinary inductive test-procedures, or allow us to forget that "probability is the guide of life."

At this point, however, a further question arises. Hume is clearly contending that experience as we have it is unable to justify the hypothesis of a transcendent deity, and hence that, on the evidence presented, the argument from design will not wash. What is not so clear is whether he thinks that the same would continue to hold good under all conceivable circumstances; whether, if new and better evidence were to appear, he would be willing to concede the argument a degree of efficacy it does not now possess. His periodic allusions to the inadequate state of our present knowledge may suggest that he would do this, and the pious hopes of a more particular revelation expressed by Philo at the end of the *Dialogues* tend to confirm this view, namely that it is the weakness and unsuitability of the premisses which vitiates the analogy between human and superhuman contrivance, and so makes the argument unacceptable. As against this, however, it would be possible to argue a case for the more radical and doubtless more purely philosophical thesis, that the very nature of the problem makes it out of the question for an *a posteriori* argument ever to succeed. Experience, it might be asserted, is always going to be finite, always subject to inductive uncertainties, always liable to misunderstanding and observational error. Hence, whatever the form it may chance to take in the future, there will always be an unbridgable gap in the attempted inference from finite to infinite, from worldly appearances of order and contrivance, however spectacular, to the existence of an intelligent almighty creator behind them. To adapt the argument of the notorious *Essay on Miracles*, it will always, perhaps, be more likely that, as finite minds, we have

made some egregious mistake in reading the evidence, than that its supposed conclusion should actually be true.

So radically sceptical an interpretation would carry Hume from a position of provisional to one of inveterate agnosticism, in which nothing in this world, or the next, is going to convince him that the design argument is valid, though he might encounter circumstances, such as the sounding of the Last Trump, for example, in which it would be the part of prudence to swallow the conclusion, notwithstanding the logical qualms that could still be felt about its legitimacy. Such qualms would by then be absurd, but not *so* absurd as those of that extravagant scepticism about ordinary matters which Hume habitually rejected. The circumstances, after all, would have to be very extraordinary—miraculous, in fact—and by definition contrary to anything that experience would have led us to expect. Belief in them is therefore bound to be reluctant, since it has to overcome the natural resistances instilled by a long habituation to the normalities of the past. Hume, if he takes this line, will therefore probably be late in arising on Judgment Day; but then, as he says, the philosophical sceptics are entitled to some indulgence and compassion for the sobriety of their opinions, even if it actually transpires that their caution was misplaced.

Some people will want to argue, of course, that as things stand, Hume is under no obligation to answer hypothetical questions about what it would take to convince him that the design argument is sound. He does, however, himself show a willingness to consider imaginary cases, in the two illustrations put forward by Cleanthes in Part III of the *Dialogues*, that of a celestial voice addressing the nations in their various languages and giving them much-needed advice, and the even more peculiar supposition of a vegetable library, in which books multiply by propagation and generate their own content. In cases such as these, so Cleanthes argues, it would be impossible to resist the inference to an extra-terrestrial or immanent intelligence behind the phenomena, and Philo, for once in a way, is said to be embarrassed. Kemp Smith dismisses the first of these illustrations as "misleading and irrelevant," and the second as "very bewildering," while denying that either of them is in any way helpful to Cleanthes' case. The fact remains, though, that Philo makes no reply to them, and

the reason for this may be that, relevant or no, they do raise an awkward question which puts him on the spot. So long as the premisses of the design argument are confined solely to natural appearances of order, mechanical regularity, symmetry, biological adaptation and the like, it is not too difficult to show that these can be produced, without intelligence, by natural selection and the repetitious operation of physical laws. Where Cleanthes' examples differ is in introducing design in a new and more literal sense of the term: the occurrence in nature of representative symbols (linguistic, as it happens) purveying ideas and information, a thing that no ordinary natural process could account for, and one that cannot, save in isolated causes, be dismissed as a mere product of chance.

Symbols imply communication, and communication implies mind. The very notion of intelligence seems to require the symbol-using ability as one of its defining conditions, and hence the clear occurrence of symbolic events in nature, in circumstances which preclude the possibility of a human agent's responsibility for them, would be manifest proof that somewhere in the chain of production a non-human intelligence was involved. The character of that intelligence, and the degree of its superiority (if any) to our own, would no doubt have to be gauged by the scale of the manifestation and the quality of the content it chose to impart. It is easy enough to imagine communications, such as the lower-case typings of Archie the Cockroach, or the traditional radio messages from outer space, which would satisfy anyone that their source was intelligent—and more so, perhaps, than we are—without in the least suggesting that it was in any way divine. That is the trouble with Cleanthes' literary vegetables and polyglot homilies from on high; they prove intelligence and power, but not the illimitable intelligence or supreme power that his argument needs in order to be even hypothetically convincing. The question they raise is whether any empirical evidence *could* be imagined, which would persuade any reasonable mind of the real existence of an infinite God.

What is wanted, evidently, is a thought-experiment which presents the principle of symbolic communication on a scale which only omnipotence could account for. Crude as it may be, the following flight of fancy will perhaps bring the matter to a head: if the stars and galaxies were to shift overnight in the firmament, rearranging

themselves so as to spell out, in various languages, such slogans as I AM THAT I AM, OR GOD IS LOVE—well, the fastidious might consider that it was all very vulgar, but would anyone lose much time in admitting that this settled the matter, and that the design argument was thereby vindicated in full? (One might imagine *other* slogans that would utterly wreck the hopes reposed in the argument; but that is not the question at issue here).

Confronted with such a demonstration, the hard-line Humean could continue, of course, to argue that, for all its colossal scale, the performance is still finite, and so cannot be evidence of more than the finite, though immense, power that is needed to achieve it. But this now seems a cavil, designed only to prove that even omnipotence is powerless against the extremer forms of sceptical intransigence. A more sensible course might be to deny that, under such conditions, the term "design" still serves as an apt description for what is evidently miracle and revelation, and no longer relies on any appeal to the customary behavior of things in nature. This saves the case so long as the phenomena are sudden and unprecedented, but will hardly do if they were themselves to figure as regular or long-standing features of the natural scene. If celestial inscriptions of this kind had first appeared, say, at the time of the Crucifixion, and were periodically altered in accordance with terrestrial circumstances, it would long since have become entirely natural to treat them as a system of messages emanating from a supreme being, and as the clearest evidence imaginable of the authenticity of a particular revelation. Hume himself could

make no complaint of the want of analogy, and this seems to show that his quarrel with the design argument is not with its use of analogical reasoning as such, but rather with the altogether too reckless deductions drawn from insufficient premisses by its 18th-century exponents. The very extravagance of the foregoing example is enough to indicate how much those premisses would require to be strengthened, if the argument that rests on them is ever to become coercive. Hume, on the evidence presented, was entirely justified in rejecting it; but there is no need, on that ground, to assume him wholly impervious to such arguments, even in theology, or unable to change his mind if he has to. To Boswell, visiting him on his death-bed, he declared it "a most unreasonable fancy that he should exist for ever," and professed no uneasiness at the thought of annihilation:

" 'Well,' said I, 'Mr Hume, I hope to triumph over you when I meet you in a future state; and remember you are not to pretend that you was joking with all this Infidelity.' 'No, No,' said he, 'But I shall have been so long there before you come that it will be nothing new.' "

The most unreasonable fancies may after all be realized, and should that happen, Hume, in his own good time, will not be unreasonably obstinate in coming to terms with them. Once he has overcome his initial incredulity, and provided that the hereafter is equipped, as it should be, with a whist-table, a backgammon board and a small stock of teasable clergymen, there seems nothing to prevent him from being entirely happy there.

IX. THE UNIVERSITY AND THE CASE FOR PREFERENTIAL TREATMENT

RICHARD WASSERSTROM

I

THE fundamental question I want to examine is whether the fact of minority membership is ever by itself a good reason for admitting a student to the university or appointing a person to a faculty position at the university. I shall argue for the plausibility of two arguments that support taking minority group membership explicitly into account for purposes of admitting students to higher education or appointing persons to the faculty. The first argument is a negative one. It is that those who deserve to be admitted to or employed by the university are not necessarily those who are the most qualified. The second argument is an affirmative one. It is that the university is benefited in very conventional and traditional terms by the presence of minority persons.

Before introducing these arguments, three points require clarification. First, I shall use the term "minority" to refer to Blacks, Chicanos, American Indians, women, and whites from a very low socioeconomic class background. I realize, of course, both that women are a numerical majority and that these categories are quite imprecise. I shall have a bit more to say about the latter point in the final portion of my paper. I do not think that the former one affects what I have to say.

Second, my thesis is different from two others with which it might be confused. I am concerned with whether minority group membership is by itself a good reason for admitting a student or appointing a person to the faculty. And that is different from either: (a) whether minority membership is by itself a *sufficient* reason for admission or appointment or (b) whether a quota system or any other particular scheme of preferential treatment is a desirable means by which to take into account this fact of the relevance of minority group membership.

And third, my ambitions are modest in another respect as well. I am concerned in this paper

with only a few of the variety of arguments that have been and that might be offered in connection with this question. I do not, for instance, say anything about the arguments based upon an appeal to principles of compensatory justice; nor do I examine whether it is constitutionally prohibited, required, or permitted to take minority membership into account; nor do I even explore many of the arguments concerning how the university is benefited by the presence of minority persons, e.g., that they serve as valuable role models. I concentrate instead upon only two arguments that seem to me potentially persuasive and somewhat less familiar.

II

The first argument I want to consider rests upon the claim that minority membership ought to be seen to be a relevant consideration just because of the benefits to be derived by minority group persons from admission to the university or appointment to the faculty. On this view, the case for a minority admission program rests upon the value to students of higher education. The first step in this argument consists in seeing that because higher education is something that is good for the students (and the good here can be conceived of in both economic and noneconomic terms), it is also good for minority group students. And the next step in the argument consists in establishing that upon some ground or other it is especially appropriate to make members of minority groups the recipients of this good, e.g., that attendance will make a substantially greater positive difference in the lives of those members of the minority group who attend the university than it will in the lives of those from the majority group who would otherwise have attended the university but who as a result cannot now do so.

Analogous arguments can be developed in support of a policy which favors the appointment

of minority group members to the faculty. For a faculty position—even more than a place within the student body—is a benefit to the holder of the position. On financial grounds alone, a professorial appointment is, relatively speaking, a good job. And when one adds to the financial picture, the security, the prestige, the freedom to decide how to spend one's time, the setting, and so on, it is a very desirable situation, indeed. Because that is so, it is certainly appropriate to argue that members of minority groups should be the recipients of this good, provided only that it can be shown that there is some reason which gives them a special claim to that good.

It is an argument such as this that many persons think to be both outrageous and wrong. And they would, I think, want to reply in something like the following fashion. The only relevant and appropriate question to be asked about admission or appointment is the question of merit: Those persons who most fully possess those qualifications for successful university work—either as a student or as a teacher—are the persons who alone deserve to be admitted or appointed. Merit is the sole consideration, especially once its criteria have been corrected for possible bias. Reparations, needs, benefits to the individual are totally and properly beside the point—a misleading diversion and distraction which at best appeals to our sympathetic emotions instead of to our intellect and our respect for the requirements of justice.

I think it important to see precisely what is and is not necessarily at stake in this objection to the relevance of minority memberships. In particular what I want to show is that what need not be at issue is any question of whether those who deserve to be admitted ought to be admitted. For I will concede, for purposes of argument, that those who deserve to be admitted or employed should be. What I take issue with is the facile assumption that merit, when measured in terms of past or projected performance, i.e., qualifications, is the sole or right test for desert. What is clear is that merit (in this sense of most qualified) is, perhaps, the single most plausible criterion for determining who will be the most successful student or the most proficient academician (although it is certainly possible to challenge on a variety of grounds the conventional conception of qualifications). However, it is an enormous leap from that truth to any

answers whatsoever to the question of who deserves to be admitted. And it is the question of who deserves to be admitted to the university that was the force behind the objection offered.

It is essential to see, too, that the problem does not exist solely because attendance at the university, either as a student or as a faculty member, bestows a number of extrinsic, valuable benefits upon the participants. The gap between who is most qualified to perform a task and who deserves to be admitted or appointed would still be present even if higher education were merely an activity of inquiry—even if the university were simply and wholly a community of intellectuals who engaged in a variety of intellectual activities for their own sake. This is so because there is simply no necessary connection between academic merit or excellence and deserving to be a member of the university community. Suppose, for instance, that there is only one tennis court in the community. Is it clear that the two best tennis players ought to be the ones permitted to use it? Why not those who were there first? Or those who will enjoy playing the most? Or those who are the worst and therefore need the greatest opportunity to practice? Or those who have the chance to play least frequently?

Now, we might, of course, have a rule that says that the best tennis players get to use the court before the others. And if we did have such a rule, then in virtue of that rule the best players would deserve the court more than the poorer ones. But that is just to push the inquiry back one stage. Is there any reason to think that good tennis players are entitled to have such a rule? Indeed, the arguments that might be given for or against such a rule are many and varied. And few if any of the arguments that might support such a rule would depend upon a connection between ability and desert.

But someone might reply that the most able students do deserve to be admitted to the university because all of their earlier schooling was a kind of competition with university admission being the prize awarded to those who came out the winners. They deserve to be admitted because that is what the rule of the competition provides. In addition, it would be unfair now to exclude them in favor of others, given the reasonable expectations they developed about the way in which their industry and their performance would be rewarded. Minority admission programs, which inevitably prefer some who are

less qualified over some who are more qualified, all possess this flaw.

There are several problems with this way of thinking about things. And the most substantial of them is that it is an empirically implausible picture of one social world. For most of what are regarded as the decisive ability characteristics for higher education have little to do with industry and a great deal to do with things over which the individual has neither control or responsibility: I.Q., home environment, socio-economic class of parents, and, of course, the quality of the primary and secondary schools attended. Since individuals do not deserve having had any of these things *vis-à-vis* other individuals, they do not, for the most part, deserve being the ones who are the most able or most qualified students. And since they do not deserve their abilities they do not in any strong sense deserve to be admitted because of their abilities.

To be sure, if there is a rule which connects, say, performance at high school with admission to college, then there is a weak sense in which those who do well at high school deserve, for that reason alone, to be admitted to college. But then, as I have said, the merits of this rule need to be explored and defended. In addition, if persons have built up or relied upon their reasonable expectations concerning performance and admission, they have a claim to be admitted on this ground as well. But it is certainly not obvious to me that these claims of desert are any stronger or more compelling than competing claims based upon the needs of or advantages to minority group members.

Having said all of this, it is also necessary to acknowledge that qualifications are relevant in at least three other respects. In the first place, it is the case that there is some minimal set of qualifications without which the benefits of participation in higher education cannot be obtained by the individuals involved. In the second place, it is also the case that the qualifications of all of the individuals within the university will affect to some degree or other the benefits obtainable to anyone within it. And finally, the qualifications of all of the individuals within the university will also affect the way the university functions *vis-à-vis* the rest of the world. The university will do some things better and some things worse, depending upon the qualifications of those who make it up. But as I have said, all these considerations only establish that qualifi-

cations, in this sense, are relevant, not that they are decisive. And as such, this is wholly consistent with the claim that minority group membership is also a relevant but not decisive consideration when it comes to matters of admission and appointment.

III

My argument so far has been essentially defensive—to weaken the appeal of the claim that those who are most qualified for that reason deserve the benefits of places within the university. For many persons, this focus upon admission or employment as a benefit may be an uncongenial way to think about the entire problem. I propose to turn, therefore, to a quite different way of looking at things, and I want now to give an argument that is designed to show how the university, itself, is benefited from the presence within it of minority group students and teachers. I believe it is an argument that ought to appeal even to those whose ways of thinking about the university are especially conservative ones. The argument is this:

Even the most traditional conception of the university sees the search for truth as one of its central functions. The addition of minority persons to the student body and the faculty will, *ceteris paribus*, increase substantially the likelihood that important truths, which would otherwise have gone undiscovered, will be discovered. Therefore, minority group membership is by itself a good reason for admitting a student or appointing a faculty member.

The crucial premise is the second one. The question, of course, is whether it is reasonable to suppose that minority group membership is connected in any significant way with the success of the university's ongoing search for truth.

One way to begin to see that it might be so connected is to see that in many, if not all, disciplines the ways in which the problems of that discipline get defined have a tremendous effect upon what eventually comes to be known and believed. Again, it is hardly a radical suggestion to observe that substantial progress in inquiry often occurs when someone succeeds in calling into question what was hitherto deemed an obvious truth or what was not even previously seen to be an assumed fact. The link between minority group membership and meaningful inquiry is forged, on this argument,

through the special likelihood that minority group persons (and especially minority group academicians) will tend to define conventional intellectual problems in new and different ways and will tend to perceive more easily unnoticed assumptions that would otherwise have gone unexamined and unchallenged.

The next step in the argument consists in making out the case that minority group persons would, within the university, tend to do so. And here, it seems to me, the place to begin is with the claim that race and sex—at least within our culture—are sufficiently central features of experience so that they affect in quite fundamental ways how it is that individuals perceive, regard, and respond to the social and physical world. It need not be maintained that race and sex are the only two significant determinants; just that they are among the more significant ones. Because they are, however, persons who are members of minority groups will tend to look at the world, and will tend to define problems quite differently from the way white, middle-class males will be inclined to.

I think that there is a good deal of truth to this claim and, consequently, a good deal of force to the argument on which it depends. In the two fields I know best, namely, law and philosophy, I think there is no question but what academicians look at a number of problems and issues differently because of the recent addition of minority group members to the student body and the faculty, and because of the questions that have been raised about the previous composition of the university. A new awareness of the many and peculiar ways in which the legal system has treated the offense of rape is one example; a more critical recognition of the United States Supreme Court's dependence in *Brown v. The Board of Education* on a conception of the superiority of white, middle-class values is another; and a reassessment of the moral problems of abortion in terms of the claims of women to control their own bodies is still a third instance. The general claim seems to me more plausible in the case of the humanities and the social sciences than it does in the natural sciences. But even in the case of some of the natural sciences, I can imagine ways in which the claim might be both true and important.

The general point is an empirical one that falls somewhere within the domain of the sociology of knowledge. It would not, therefore, be

appropriate, even if I had the competence to do so, for me to try to marshal in a systematic way the relevant evidence bearing on its truth or falsity. Nonetheless, it does seem to me that there are several aspects of the thesis that can profitably be explored both to introduce some relatively neglected methodological considerations and to indicate in a general, informal way one reason to believe the claim to be a more plausible and important one than it is often taken to be.

In the first place, it may be difficult to get the thesis taken seriously because part of the characteristic, white, middle-class point of view consists in the belief that reasonably well-educated, well-intentioned white, middle-class males possess the capacity to view both social and natural phenomena in a detached, objective, non-distorted fashion. Sometimes, of course, this view derives simply from a more general one concerning the comparable objectivity of all persons similarly situated. More often and more interestingly, however, the view is one that white, middle-class males tend to hold exclusively about themselves. That is to say, they are quite ready and eager to acknowledge that others—members of various racial groups, or women, or persons of other social classes—do look at the world, approach problems, define issues, etc., through particular, nonobjective points of view. But while this is something others do, it is not something that they do; for they possess the capacity and the detachment to look at things fairly, comprehensively, and completely—in short, to view things as they really are.

Such an attitude is reflected in the following tale. There was a case once in which the plaintiff was black and the defendant was white. It was a contract case having nothing very much to do with race. The judge who was assigned to hear the case was himself black. Just before the trial was about to begin, the attorney for the white defendant, who was himself white, asked the judge if he could talk with him for a moment in chambers. The judge agreed and they repaired to the judge's chambers. The lawyer then asked the judge if he, the judge, didn't think that it would be better for all concerned if the judge disqualified himself from the case because it would be hard for him to be fair, given the fact that the plaintiff was black, too.

Now I do not think this is an especially instructive way to show that lots of whites believe

that no black judge could be fair in judging a case between a black party and a white party. But I do think this is a useful way to make the other point—that an assumption of impartiality and objectivity is very often a part of the white, male, middle-class view of itself. I am interested, that is, in the lawyer's ready, unreflective recourse to the typical judge, the white judge, as the person who would be best able to look at the case in the way that justice and neutrality required.

If I am right that this is often a part of the white, middle-class, male outlook, then this creates a meta-problem at the outset. For among the assumptions of the white, male, middle-class point of view that it will be hardest to call into question is the assumption of the typical detachment, neutrality, and objectivity of the white, middle-class males in question—and they are the ones who in overwhelming numbers do make up the faculties of our universities.

In the second place, it is important to be clearer about how extensive the thesis is. Some might be inclined to accept it only because they interpret it in an unduly restrictive fashion; as the claim that the race or sex or class of the persons involved will make a substantial difference in the way issues get thought about only where the matters under consideration themselves directly involve race, sex, or class. That is to say, the thesis might be deemed plausible only in circumstances where one's interest as a member of a minority group is somehow directly relevant. I think, however, that this is an unnecessarily cramped conception of the general point (and, in fact, is probably a reflection of the white, male, middle-class point of view delineated above). A more ambitious interpretation of the thesis certainly seems to me to be both possible and plausible; it regards the race, sex, or class of the relevant individuals as affecting their points of view in respect to many matters that are not directly racial, sexual, or class-related.

Once again, an example drawn from the law can be instructive. Consider the question of whether it is important for different minorities to be present on juries, and if so why it is important. One view, which I take to be both dominant and overly simplistic, thinks that minority representation is important only if, or because, all jurors will tend to act upon straightforward prejudices as jurors. White jurors will, for instance, tend to vote against black defendants and black jurors

will tend to vote in favor of them—irrespective, to some degree, of the evidence. On this view minority representation makes sense as the way to cancel the unfair voting of other interests. I do not believe this is a very sensible way to think about the behavior of jurors, and I do not, therefore, believe that it makes very convincing the claim that a black person cannot have a fair trial at the hands of an all white jury (although I think that is a convincing claim).

But the point is that this is not the only way to think about things. A black person may not be able to get a fair trial from a group of twelve white jurors, even though they are disposed to be fair and impartial, because the whites may simply bring with them into the jury box a general point of view which affects in very fundamental respects the way they look at the facts.

I was, for example, once involved in a case in which a young black male was charged with having thrown, without provocation, an empty coke bottle at a passing police car. As my client told it to me, however, nothing of the sort had occurred. Instead, he had been standing on the corner selling copies of the Black Panther newspaper when the police car drove by. The police slowed down and said: "Nigger, if we see you on the street with those papers again we will kill you." My client replied: "You @\$\$!\$ts can't scare me." The police stopped the car, chased and caught the youth, and took him to jail. They then charged him with throwing an object at a vehicle, a felony which carries in California a sentence of up to five years in the penitentiary.

At the preliminary hearing, the officers, when they told their story, were insistent that they had made no mistake—they had seen the defendant clearly, he had thrown the bottle at their car, and he had done so with a pitching motion with his right hand. The magistrate held the defendant over for trial. Fortunately for the defendant he was left-handed, markedly so, and we also had a witness who had seen the incident and who was sure that no bottle had been thrown. The defendant had a jury trial before twelve white jurors. The police told the same story they told at the preliminary hearing, and we then put on our eye witness together with a doctor who explained to the jury how he could tell that the defendant was left-handed. The jury then retired to deliberate, and in a bit more

than seven hours it returned with a verdict of acquittal.

Now someone might think that because the jury found the defendant not guilty, this shows how fair the jury was. But I do not think so. I think the interesting thing is that it took the jury more than seven hours to acquit the defendant. I think the reason why it took so long is because the jurors were white and middle class. And I do not at all mean to suggest that they wanted to convict the defendant because they were white and he was black. Rather, I think that their general white, middle-class point of view made it hard for them to reach the correct decision. My hypothesis is that the problem was rooted in the difficulty the jurors had in convincing themselves that the police officers had deliberately fabricated a story about the defendant. I know that this was the problem I had with the case for a very long time. For there was simply nothing whatsoever in my experience with police that made it easy for me seriously to entertain the belief that they would be prepared to lie in so flagrant a fashion. I suspect, however, that the typical black experience is quite different. Many black persons, just because they are black, have had encounters with police in which the police were indifferent to the truth of what they said. And even more black persons have known a friend, relative, or associate who has had such an experience. If this is so, then one probable result is that a part of the typical black point of view consists in a quite different presumption about the credibility of police than that which is contained within the typical white, middle-class point of view. Thus, on this analy-

sis the case for representational juries rests not upon the counteracting-prejudice theory but upon the necessity of having different points of view of this sort reflected in the jury's deliberations, if trials are to be genuinely fair.

It is this same kind of argument, I believe, that conceivably holds in respect to substantial minority presence in the student body and on the faculty. For if there are these distinctive points of view and if these points of view are typically connected with minority membership, then the case for the relevance of minority identity is in part the case for a useful and valuable type of intellectual pluralism which aids the university in the fulfillment of one of its most conventional and most important functions—the pursuit of knowledge.

The argument may not, of course, be a good one because I am mistaken about the facts. It may be that it is not possible to formulate a coherent and meaningful account of the substance of these different points of view. Or it may be, instead, that these different points of view are not even typically connected with identifiable minority groups. These seem to me to be two of the more central questions that require further exploration. But it is also important to see that if the facts are as I suspect them to be then an additional argument is generated. For on this account it is the university that is benefited (in terms of its most conventional end) by the presence of minority persons—yet in respects not always readily perceived by those who most often identify the university exclusively in terms of this same conventional end.

University of California, Los Angeles

Received August 21, 1975

BOOKS RECEIVED

- Addis, Laird, *The Logic of Society* (Minneapolis: University of Minnesota Press, 1975), 226 pp., \$12.50.
- Ajdukiewicz, Kazimierz, *Pragmatic Logic* (Dordrecht: D. Reidel Publishing Co., 1974), 460 pp.
- al-Tabataba'i, Muhammad Husayn, *Shi'ite Islam*, tr. by Seyyed Hossein Nasr (Albany: State University of New York Press, 1974.) \$10.00
- Apostel, Leo, *Matiere et Forme I & II* (Gand, Belgium: Centre National de Recherches de Logique, 1974), 454 & 452 pp.
- Beauchamp, Tom L. (ed.), *Ethics and Public Policy* (Englewood Cliffs: Prentice-Hall, Inc., 1975), 442 pp.
- Best, David, *Expression in Movement & the Arts* (London: Lepus Books, 1974), 203 pp.
- Blackwood, R. T. and A. L. Herman (eds.), *Problems in Philosophy West & East* (Englewood Cliffs: Prentice-Hall, Inc., 1975), 474 pp.
- Blumenthal, Albert, *Moral Responsibility: Mankind's Greatest Need* (Santa Ana, Calif.: Rayline Press, 1975), 474 pp.
- Bontempo, Charles J. and S. Jack Odell (eds.), *The Owl of Minerva* (New York: McGraw-Hill, 1975), 264 pp., \$10.00.
- Bosley, Richard, *Aspects of Aristotle's Logic* (Atlantic Highlands, N.J.: Humanities Press, 1975), 137 pp. \$10.00
- Dumain, Harold J., *A Synthesis of Philosophy* (New York: Philosophical Library, 1975), 185 pp., \$6.00.
- Gill, Richard and Ernest Sherman (eds.), *The Fabric of Existentialism* (Englewood Cliffs: Prentice-Hall, Inc., 1973), 640 pp.
- Gonzalez, Jose Lorenzo, *Origen Biologico de los Saberes Logico-Matematicos* (Salamanca: Universidad Pontificia, 1975), 184 pp.
- Grzegorzczuk, Andrzej, *An Outline of Mathematical Logic* (Dordrecht: D. Reidel Publishing Co., 1974), 596 pp.
- Hook, Sidney, Paul Kurtz, and Miro Todorovich (eds.), *The Philosophy of the Curriculum: The Need for General Education* (Buffalo: Prometheus Books, 1975), 281 pp. \$10.95
- Hourani, George F., *Essays on Islamic Philosophy and Science* (Albany: State University of New York Press, 1975), 261 pp.
- Johnson, Oliver A., *The Problem of Knowledge* (The Hague: Martinus Nijhoff, 1974), 131 pp.
- Kohl, Marvin (ed.), *Beneficent Euthanasia* (Buffalo: Prometheus Books, 1975), 255 pp. \$10.95 & \$4.95
- Kristeller, Paul Oskar, *Medieval Aspects of Renaissance Learning* (Durham: Duke University Press, 1974), ed. by Edward P. Mahoney, 175 pp. \$7.50
- Lentricchia, Frank, *Robert Frost: Modern Poetics and the Landscapes of Self* (Durham: Duke University Press, 1975), 200 pp., \$8.75.
- Miller, Dickinson S., *Philosophical Analysis and Human Welfare* (Dordrecht: D. Reidel Publishing Co., 1975), ed. by Loyd D. Easton, 333 pp.
- Pullias, Earl V., *A Common Sense Philosophy for Modern Man* (New York: Philosophical Library, 1975), 209 pp. \$8.75
- Purtill, Richard L., *Philosophically Speaking* (Englewood Cliffs: Prentice-Hall, 1975), 226 pp.
- Schacht, Richard, *Hegel and After: Studies in Continental Philosophy Between Kant and Sartre* (Pittsburgh: University of Pittsburgh Press, 1975), 297 pp., \$15.95, \$3.95.
- Sherif, Mohamed Ahmed, *Ghazali's Theory of Virtue* (Albany: State University of New York Press, 1975), 205 pp.
- Sosa, Ernest (ed.), *Causation and Conditionals* (Oxford: Oxford University Press, 1975), 202 pp.
- Stegmüller, Wolfgang, *Hauptströmungen der Gegenwartsphilosophie, I & II* (Stuttgart: Alfred Kröner Verlag, 1975), 730 & 566 pp.
- Straus, Erwin W. (ed.), *Language and Language Disturbances* (Atlantic Highlands, N.J.: Humanities Press, 1974), 267 pp. \$12.50
- Stuewer, Roger H. (ed.), *Historical and Philosophical Perspectives of Science* (Minneapolis: University of Minnesota Press, 1970), 384 pp., \$11.50.
- Voegelin, Eric, *From Enlightenment to Revolution*, ed. by John H. Halliwell (Durham: Duke University Press, 1975), 307 pp. \$12.75
- Welsh, Paul (ed.), *Fact, Value, & Perception: Essays in Honor of Charles A. Baylis* (Durham: Duke University Press, 1975), 174 pp. \$8.75
- Zakopoulos, Athenagoras N., *Plato on Man* (New York: Philosophical Library, 1975), 142 pp., \$7.50.
- Zellner, Harold M. (ed.), *Assassination* (Cambridge, Mass.: Schenkman Publishing Co., 1974), 111 pp.

AMERICAN PHILOSOPHICAL QUARTERLY MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

In March 1976 we shall be publishing **NEGATION AND NON-BEING** by Richard Gale, which is the tenth in this impressive Monograph series. The price is \$8.00 (£4.00) to institutional subscribers and to non-subscribers; individual subscribers, however, will continue to receive the Monograph as part of their subscription. In past years we were able to supply Monographs to all subscribers to the Quarterly, but this has for the time being become impossible. If your library is a subscriber to the Quarterly, please check whether it is now subscribing also to the Monograph series. It is only in the Monographs that the larger themes of modern philosophy can receive extended treatment, and those who subscribe only to the Quarterly are missing an essential part of the coverage of philosophy which is provided by the journal as a whole.

To enable you to judge for yourself the importance of this series, a complete list of the Monographs is given overleaf. The last issue supplied to institutions on subscription was No. 7 (1973). If you find that Nos. 8 (1974) and 9 (1975) are still missing from your files, we shall be glad to supply them at the quoted prices, and to enter your standing order for all future issues of the series.

PRINTED IN ENGLAND

by William Clowes & Sons Limited, London, Beccles and Colchester

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

- No. 1. **STUDIES IN MORAL PHILOSOPHY.** *Contents:* Kai Nielson, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supercorroboration and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$8.00 (£4.00).
- No. 2. **STUDIES IN LOGICAL THEORY.** *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Höderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$8.00 (£4.00).
- No. 3. **STUDIES IN THE PHILOSOPHY OF SCIENCE.** *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$8.00 (£4.00).
- No. 4. **STUDIES IN THE THEORY OF KNOWLEDGE.** *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, \$8.00 (£4.00).
- No. 5. **DESCARTES' PHILOSOPHY OF NATURE** by James Collins. 1971, \$8.00 (£4.00).
- No. 6. **STUDIES IN THE PHILOSOPHY OF MIND.** *Contents:* Harold Brown, "Perception and Meaning"; Haskell Fain and A. P. Griffiths, "On Falsely Believing That One Doesn't Know"; O. H. Green, "Emotions and Belief"; Donald F. Henze, "Descartes on Other Minds"; Moreland Perkins, "Meaning and Feeling"; and Alan R. White, "What We Believe." 1972, \$8.00 (£4.00).
- No. 7. **STUDIES IN ETHICS.** *Contents:* Charles V. Blatz, "Mitigating and Meliorating Defenses"; Claudia Card, "Retributive Penal Liability"; Norman O. Dahl, "Is Mill's Hedonism Inconsistent?"; R. L. Franklin, "Recent Work on Ethical Naturalism"; James D. Wallace, "Cowardice and Courage"; A. D. Woozley, "Injustice." 1973, \$8.00 (£4.00).
- No. 8. **STUDIES IN MODALITY** by Nicholas Rescher, with the collaboration of Ruth Manor, Arnold vander Nat, and Zane Parks. 1974, \$8.00 (£4.00).
- No. 9. **STUDIES IN EPISTEMOLOGY.** *Contents:* Dennis M. Ahern, "Hume on the Evidential Impossibility of Miracles"; Robert Almeder, "Fallibilism and the Ultimate Irreversible Opinion"; Jonathan Bennet, "Stimulus, Response, Meaning"; Peter A. French, "'Seeing' and 'Seeing that', 'Observing' and 'Observing that'"; Margaret Gilbert and Fred R. Berger, "On an Argument for the Impossibility of Prediction in the Social Sciences"; David K. Glidden, "Protagorean Relativism and Cyrenaics"; Mark Pastin, "Modest Foundationalism and Self-Warrant"; E. E. Sleinis, "Normal Conditions and Perceivers". 1975, \$8.00 (£4.00).

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Alvin Plantinga
Jerome Schneewind
George A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 13/NUMBER 4

OCTOBER, 1976

CONTENTS

- | | | | |
|---|-----|--|-----|
| I. ZELIG HARRIS: <i>A Theory of Language Structure</i> | 237 | V. DOUGLAS ODEGARD: <i>Parasitical Reference and Paradox</i> | 295 |
| II. WILLIAM P. ALSTON: <i>Self-Warrant: A Neglected Form of Privileged Access</i> | 257 | | |
| III. JAMES R. FLYNN: <i>The Realm of the Moral</i> | 273 | VI. JOHN A. BARKER: <i>What You Don't Know Won't Hurt You?</i> | 303 |
| IV. ANDREW WARD: <i>Direct and Indirect Realism</i> | 287 | VII. CLYDE LEE MILLER: <i>Ironic or Not?</i> | 309 |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes" (short or long). Moreover the editor will only consider papers whose authors certify that while under consideration with us, they will not be submitted elsewhere.

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The annual subscription for 1977 will be \$28.00 for institutions, \$12.00 to individuals. Single issues will be \$8.00 to all purchasers. For Monograph prices, see below. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The complete series is listed on the back cover. In 1977 these will be available at \$9.00 to institutions and non-subscribers, \$4.50 to individual subscribers. Please note that it is no longer possible to supply the monographs free of charge to individual subscribers.



I. A THEORY OF LANGUAGE STRUCTURE

ZELLIG HARRIS

THIS paper presents a theory of the structure and information of sentences, which consists of three fundamental relations: the ordered entry of words into the making of a sentence; inequalities of likelihood of occurrence of a particular *n*th entering word in a sentence in respect to the choices of words entering immediately before the *n*th; and the reduction in shape of certain words which have, upon their entry into a particular sentence, exceptionally high likelihood of occurrence, i.e. low amount of information. The first two relations determine the structure of a base subset of sentences which have a transparent grammatical structure, and contain little or no grammatical restrictions, and carry all the substantive information carried in any sentence of the language. The last relation is in general optional, and does not materially change the information in a sentence. It introduces a secondary grammatical complexity into sentences, and introduces restrictions, subsets, and grammatical rules into the language. From the base sentences all the remaining sentences of the language can be obtained by a direct application of the stated set of reductions, in a manner suitable to the constructing and, with certain short-cuts, to analyzing of sentences.

The adequacy of these fundamental relations to characterize the sentences of the language, and the simple nature of the relations themselves, suggests that the structure of language is relatable to the conditions of language's existence without appeal to some independently existing structuralism. It may be philosophically relevant that the structure and semantic relations of sentences can be obtained from this model, and that the sentence types which are involved in various philosophical issues can be identified in particular ways as cases of these general structural relations.¹

I. OPERATORS

1. Entry Order

The theory starts from a relation between word-occurrences in a sentence, when the juxtaposition of one word to another is sufficient to constitute a sentence. It is similar to the relation noted in categorial grammar;² but the use made of it here is different. The relation is clearest when we consider that many sentences contain sentences, in altered or unaltered form, together with word-sequences that are not themselves sentences. For example, the sentences *Mary phoned*, *John arrived* are contained in: *Mary's phoning ceased*,³ *John denies that Mary phoned*, *Mary's phoning entailed John's arriving*, *Mary phoned and John arrived*. To treat this relation, we introduce a term: If the characterization of *B* involves reference to *A*, but the characterization of *A* does not involve reference to *B*, *A* is called descriptively prior to *B*, and *B* descriptively later than *A*. In the sentences above, *Mary phoned* is descriptively prior to *Mary's phoning ceased*. The only status that *ceased* has in respect to the set of sentences is that it constitutes the increment between certain descriptively prior sentences such as *Mary phoned* and corresponding later sentences such as *Mary's phoning ceased*. We therefore say that in the latter sentence, *ceased* enters after the included sentence *Mary phoned*, and that *ceased* is the operator (or later entry) on *Mary phoned*, with *Mary phoned* as its argument (or prior entry). Similarly, *entailed* and *and* are operators on two arguments: on the sentence-pair *Mary phoned*, *John arrived*.

We next consider sentences which contain no sentence as a proper part of them, and where we cannot say that one part is descriptively prior to another. In *Mary phoned*, the word *phoned* has the same morphology as *ceased*, *entailed* (though not as *and*), i.e. it receives tenses and certain other affixes

¹ The writer's deepest thanks are due to Henry Hiz for major criticisms of this paper, and to Danuta Hiz for many valuable comments on precision of the analyses and formulations.

² See Kazimierz Ajdukiewicz, "Die syntaktische Konnexität," *Studia Philosophica* 1.1-27 (Syntactic Connexion, pp. 207-231, in Storrs McCall, (ed.), *Polish Logic 1920-1939*, Oxford 1967); Joachim Lambek, "The mathematics of sentence structure," *American Mathematical Monthly*, vol. 65 (1958), pp. 154-170; Henry Hiz, "Grammar Logicism," *The Monist*, vol. 51 (1967), pp. 110-127; "Computable and uncomputable elements of syntax," *Logic, Methodology and Philosophy of Science III*, ed. by B. van Rootselaar and J. F. Staal (Amsterdam, 1968), pp. 239-254.

³ The alteration by ...s ...ing, and the insertion of *that* here, are introduced as indicators of argumenthood in 1.1, end.

ment is a pair or triple of argument-words impose a preposition as argument-mark on their non-first argument-words: *John relies on Frank*, *John attributes this to Frank*, *John attributes this to her having phoned*. In some languages the elementary arguments (sect. 2) of an operator receive suffixes (called case-endings) that indicate their argument order: nominative for first argument-word, dative for second or third argument-word of certain operators, etc.

2. Argument-Requirement Sets

We next consider the possibility of classifying words in respect to their entry into discourses. If we ask about the likelihood of occurrence of individual words, we find that each word, in entering as operator, has a unique gradation in the likelihood of its occurrence, in respect to the various words which can appear in a given argument position under it. Despite the uniqueness and gradation, a classification of words in terms of likelihood is achievable if we distinguish zero likelihood from non-zero. Thus, *slept* has in its lone argument position *John*, *dog* (*John slept*, *The dog slept*), less frequently *fish*, still less *tree* (as in *The trees slept every winter until the sap began to flow*), more rarely but not impossibly *bacteria*, *virus*, *rock* (as in *The bacteria slept because of the low temperature*, *These rocks slept here through the ages*), but presumably never (except in linguistic discussion) *because*, *go* (as in *John's returning because we phoned slept*, *John's going slept*). And *entails* has in its two argument positions the pairs *return*, *leave* and *win*, *move* (as in *My winning entails our moving*) but not *he*, *me* (*He entails me*). We can now define as the *n*th place argument-requirement (or argument-demand) of a given operator the set of all words which have non-zero likelihoods of occurring in the *n*th argument position of that operator. Thus the first-place argument-demand of *slept* includes *John*, *dog*, *fish*, *tree*, *bacteria*, *virus*, *rock*, but not *because*, *go*; the second-place argument-demand of *entails* includes *leave*, *move*, but not *me*. The whole argument-demand of an operator which has *n* arguments is the set of ordered *n*-tuples of its arguments in the *n* positions: for *entails*, it includes *return*, *leave* and *win*, *move*, but not *he*, *me*.

We now find that only two word-sets appear in argument-demands, for any of the operators: The whole set of operators, *O*, on whatever arguments;

and the set of words, *N*, which are not operators on any arguments.⁵ Thus, the argument requirement of *slept* is the set of all words which are not themselves operators (except as in fn. 5), and the argument requirement of *entails* is the set of all pairs of operators.

To show that argument demands are indeed only *N* or *O*:

(1) There are operators which have both *N* and *O* in one of their argument positions (e.g. *John caused this* and *John's phoning caused this*); and there are words which have arguments in certain occurrences but not in others (e.g. *The paper tore*, *He papered the walls*). In such cases, one form can be shown to consist of the other plus reductions (section II). There are also words which have arguments in certain occurrences and not in others, without satisfying the conditions for one form to be a reduction of the other. In English, these cases are not regular, and can best be understood as a word appearing in two sets: e.g. *The rock fell*, *The boat rocked*; *He ate a prune*, *They prune the trees*.

(2) Operators are not restricted to occurring on proper subsets of *N* or *O*, with the possible exception in one language or another of a few restricted subsets which would have to be listed. Thus *slept* is not confined to arguments naming animals or the like. Its use with other arguments is not a matter merely of metaphor, but also of extended and marginal meanings: a horticulturalist may consider that his trees indeed sleep the winter and need a particular amount of sleep. In addition, there is the use of words in altered perceptions of the world, whether serious or in speculative and imaginative writing: a story might have a house speaking to its inhabitants, going to sleep, etc. True, the house is then seen as animated, but this does not change the fact that *The house slept* then occurs, and not even in a metaphoric sense. There is no member of *N* which can be assured of not occurring as argument of *slept*. Even *The vacuum slept* might occur as a sentence, say in a child's cosmology book, in a way that *Go the a* would not.

From this there follows: There must exist a certain subset of words, or word occurrences in sentences, that have no arguments, since the first other word (outside this subset) to enter a sentence can do so only if an argument-less word occurs in the sen-

⁵ There are in addition two listable sets of words in each language: (1) words which are shown in the present theory to be reduced forms of particular operators or of word-sequences constructed by the operator-argument relation described here: e.g. *the*; (2) words which do not enter into grammatical combinations, i.e. are arguments only of metalinguistic operators (sect. 4): e.g. *hello*, *ouch* (in *He said: Hello*). These last are a limiting case of the sentence-making entry-order in that they can be considered as constituting a sentence by themselves.

tence (even if that word is later zeroed, sects. 6-8). These argument-less words are the elementary arguments, N , such as *Mary*, *rock*, *virus*. In addition, if a language has any words other than elementary arguments—and it must if it is to have sentences—it must have some words, or word-occurrences, whose arguments are only N , since a word any of whose arguments is itself an operator could not enter a sentence whose only prior entries are N . These are the elementary operators: O_n (*slept*), O_{nn} (*ate*), etc. In addition, a language can have also words, or word-occurrences, some of whose arguments are themselves operators. These are the non-elementary operators: O_o (*continued*), O_{no} (*denied*), etc. O_{bo} indicates a word whose argument-requirement is the ordered pair B, C .

That N and O , without specifying subsets of them, suffice to characterize the argument-sets of all or almost all operators is supported by the fact that the set of argument-demands of all words is the set of all permutations of N and O from length zero to length 3 or 4. That is, there are words N with zero argument, words O_n with one N argument, O_{nn} with two (e.g. *ate* in *John ate fish*), O_{nnn} with three N (e.g. *put* in *John put the book on the table*, but not *John put the book*), perhaps O_{nnnn} (e.g. . . . *interpose . . . between . . . and . . .*), O_o with one operator as argument, O_{oo} with two operators as arguments (e.g. *entails*), O_{no} with N and an operator as ordered arguments (e.g. *denied*), O_{on} with the reverse order (e.g. *amused* in *John's phoning amused me*), O_{nno} (e.g. *told* in *John told Mary of Frank's phoning*), O_{noo} (e.g. *attribute* in *I attribute her leaving to his phoning*). These prove adequate as syntactic categories for the language.

Although the general conditions for operator entry are as given above, various languages may have a finite (and reasonably small) number of restrictions applying to particular subsets of operators. A subset of operators may be so expected in each sentence as to be virtually required (e.g. the *before*, *after* source of tense). In a particular subset, e.g. *continue*, *cease*, etc., the occurrence of one operator may preclude any other one—or itself iterated—from operating on it. However, strong as such restrictions may be in the overt grammar of affixes, etc., they generally turn out to be only selectional (sect. 5) and not rigid in the underlying word-entry operators.

Beginning merely with ordered word-entry to make a sentence, a fair amount of structure has thus been obtained. In summary: The entry of words into sentences depends upon the word's

argument-demand; the arguments are identified only by whether they themselves have arguments or do not have arguments, or are themselves zero, and not by belonging to particular subsets of these, or by other properties; and every resultant of an operator with arguments that satisfy its argument-demand is a discourse (and a sentence). From this it follows that the set of discourses is the set of all sequences satisfying the argument-demand relation. For the structural analysis, it is not necessary to identify the objects in the sequence as words, let alone words of a particular type, since their condition of entry depends entirely on their relation to objects defined in turn by their conditions of entry.

3. *Unrestrictedness*

We have seen (in sect. 2(2)) that, except for special cases (below), operators are not restricted to particular subsets of the N and O arguments. We will now see that we do not need to admit any subsets of operators distinguished by position in respect to their arguments. That is to say, given an operator set O_{bo} with argument-requirement BC , it will not be the case that one proper subset, O'_{bo} , makes sentences only by occurring in one position with respect to BC (e.g. $B O'_{bo} C$), while another, O''_{bo} , occurs only in another position (e.g. $BC O''_{bo}$). If this were the case, operators would have to be characterized not only by their argument-demand but also by their position in respect to their arguments, and one might think that each position indicated something different than what would be indicated by the operator words themselves. Many languages seem to have such position-subsets of operators; but it can be shown that there is one unrestrictive position in which there can occur all operators (or transformational equivalents of them in a manner indicated below) having a given argument-demand, and that any operators having that argument-demand which are restricted to other positions can be characterized as reductions (by section II) from particular operators in the unrestrictive position. The restriction becomes then a matter of subsets of operators receiving a reduction, rather than of subsets of operators having different relations to their arguments.

To show this requires a rather involved discussion. We begin with two properties of word-occurrence likelihoods: (1) For each operator word, some words in its argument-demand are more likely to occur as its arguments than are other words in the set; this likelihood-gradation differs from one operator to another, and is related to its

meaning, as in the more and less likely arguments of *slept*. Similarly for arguments, in respect to operators on them. (2) For certain O_{∞} (e.g. *cause*), what is most likely in the arguments is that their second argument or some argument further down under it, even if zeroed, should be the same word as their first argument or some argument under it in turn. Thus in any two sentences under *cause*, word-repetition is usually present or implicit (i.e. zeroed by sect. 8): In *The loss of his umbrella caused him to buy a new umbrella*, *umbrella* occurs in both operands of *caused* (i.e. in the two sentences operated on by *caused*). In *The threat of rain caused him to buy an umbrella* there is no repetition; but among the implicit sentences, including dictionary-definitions, which could have been joined to *rain* and have been zeroed are sentences such as *umbrellas protect from rain*. In contrast, *Its being Tuesday caused him to buy an umbrella* is unlikely, as containing no overt or zeroed repetition.⁶

We now raise the question of how it is that every language can express almost any information, though each has a limited vocabulary. To see how this capacity of language arises, consider the following: Let C_1 be a particular word, e.g. *house*, occurring as operator or as one of the arguments in all sentences of a set of sentences $A B C_1$. Some word pairs in the word sets A, B are most unlikely with C_1 ; the other pairs, which are not especially unlikely, will be marked $A_1 B_1$. Thus *John built*, *Fire burned* are in $A_1 B_1$ (yielding *John built a house*, *Fire burned a house*), but *John seated*, *Fire melted* are not (**John seated a house*, **Fire melted a house*). To each sentence of the set $A B C_1$, we now adjoin $O_{\infty} X Y Z$, where O_{∞} is an operator of the type in (2) above and $X Y Z$ is a particular individual sentence which is second argument of that O_{∞} . In the resultant set of sentences $A B C_1 O_{\infty} X Y Z$ (where all symbols except A, B indicate particular words), the members of A, B which are not especially unlikely are not the same as those indicated by $A_1 B_1$. Thus, let $O_{\infty} X Y Z$ be (a) *which is made of snow* where O_{∞} = semicolon plus *wh-* (relative

clause) and $X Y Z$ = *house is made of snow*, with *-ich* as pronoun for *house*; or, with reductions, (b) *made of snow*. The more likely A, B pairs here, marked $A_1 B_1$, include *John built*, *Fire melted* (as in *John built a house made of snow*, *Fire melted a house built of snow*), while other pairs such as *John seated*, *Fire burned* are not in $A_1 B_1$ (**John seated a house made of snow*, **Fire burned a house made of snow*). The difference between $A_1 B_1$ and A, B arises because under the O_{∞} there exists greater likelihood for those $A B$ words which are semantically related to the fixed $X Y Z$ words. Thus the particular word-sequence $C_1 O_{\infty} X Y Z$ determines a different likelihood-gradation in $A B$ than does C_1 alone: $A_1 B_1$ instead of $A_1 B_1$. Indeed, by (1) above, $C_1 O_{\infty} X Y Z$ acts like a new member occupying the C position in $A B C$. In this way, grammar creates word-sequences that do the work of an indefinitely expandable vocabulary. Today, English vocabulary contains a word *igloo*, borrowed from Eskimo, which determines approximately the same operators on it (and co-arguments under those operators) as does *house made of snow*. But the possibility of having the likelihood-gradation which is characteristic for operators on *igloo* (and so the meaning of that word) did not require the existence of that word in English; it was available for *house made of snow*.

To return now to the restrictive positions of operators. Most operators O_0 on a sentence, hence on an operator, occur after their argument, e.g. *continued*, is an *art*, in *John's solving puzzles continued*, *John's solving puzzles is an art*; these will be indicated here by O'_0 . Only a few, O''_0 , occur at an interior point of their operand, between their immediate argument and its argument in turn: e.g. the auxiliaries *can*, *may*, etc., and also such verbs as *continued*, in *John can solve puzzles*, *John may solve puzzles*, *John continued solving puzzles*. Following the discussion above, for every O'_0 with its particular likelihood-gradation on its operand sentences, it is always possible to find a paraphrasing $O''_0 O_{\infty} X Y Z$ such that the likelihood-gradation which it imposes on its operand sentences differs from the

⁶ Zellig Harris, *Mathematical Structures of Language*, Interscience Tracts in Mathematics 21 (New York, 1968), pp. 131-135. The finding is that in every occurrence of *cause* between two sentences which is considered likely (or, in terms of hearer's response, acceptable) either such repetition is seen, or there are zeroable (implicit) sentences, such as definitions or other likely sentences, adjoined to one of the arguments, which supply the repetition. Such a sentence would be *The threat of rain, for which umbrellas are a protection, caused him to buy an umbrella*. Definitional sentences joined to an argument by semicolon plus *wh-* (relative clause) are zeroable (*ibid.*, pp. 78-83, 137-138). *Its being Tuesday caused him to buy an umbrella* is felt as there being no semantic connection which would explain the casual relation; but grammatically what we see is the absence of overt or grammatically-zeroable repetition. One could explain the causal relation by adjoining to *Tuesday* the sentence *and rain was predicted for Tuesday*; but this would not be zeroable by the reductions of section II below. Had the sentence been *Its being Tuesday, with rain predicted for Tuesday, caused him to buy an umbrella* (with zeroed *which protects against rain* at the end), the sentence would seem perfectly likely, with zeroed repetition of *rain*.

likelihood-gradation imposed by O'_0 and approaches that of the given O'_0 as closely as we wish. The only question that might arise is if the O'_0 have a restrictive grammatical property (aside from their individual likelihood-gradations) which does not hold for the O'_0 . Indeed, the O'_0 carry a special reference to the first argument of their operand sentence: *continued* refers to *John* more directly in *John continued to swim* than in *John's swimming continued*. However, the fact that we can adjoin to a sentence metalinguistic references to any part of that sentence (sects. 4 and 16) assures the possibility of adjoining to the selected O'_0 O_{oo} $X Y Z$ a metalinguistic statement giving it the O'_0 property. Thus for *continued* in O'_0 we may say *was continued by the subject*; for *can* in O'_0 we may say *is a capability of the subject's* (*John can swim*, *John's swimming is a capability of the subject's*). In contrast, we cannot construct in the more restrictive O'_0 position a paraphrase of each O'_0 , because no metalinguistic statement on O'_0 words can remove from them the special property (the subject reference) which adheres to being in the O'_0 position.

It follows that for each argument-demand, there is a position, which in English is found to be always after the first argument, in which every operator having that argument-requirement either itself occurs or else has a paraphrase which is equivalent to it in likelihood-gradation and metalinguistic properties. It will be seen below that in English the equivalents in the post-first-argument position satisfy the conditions for being the base form in respect to which the restricted operators are reductions (transformations). Thus we can say that the post-first-argument position is unrestricted, containing all operators; and that for some operators there take place reductions (e.g. to *can*) which send them to another position, the reductions being restricted to particular members of the operator set.

We have thus arrived at a system of operators which are not restricted, in respect either to subsets of their arguments or to the position of the operators. There may nevertheless remain, in a particular language, certain subsets of operators which are restricted. Thus, for a particular set of operators which are restricted in their position (such as the English auxiliaries), or in their ability to occur under certain further operators (such as the aspectually-selective verb stems in Slavic languages), it may be difficult to find satisfactory unrestricted paraphrases; or these paraphrases may be too complex or metalinguistic. Also, some sets of operators may be such that even in their most unrestricted

paraphrase they cannot iterate with themselves or with each other. In all these cases, we are dealing with specifiable subsets of operators whose members have (approximately) identical or complementary inequalities of likelihood of occurrence in respect to another specifiable subset of operators on them (e.g. certain operators describing number, durativity); and the members of the subsets would have to be listed.

4. Metatextual Operators

There are a few operators or operator-argument combinations which differ from all others in that they name the entry-order of a word in the sentence on which these operators are acting: that is, they refer to a word within their own argument, and will be termed "metatextual" (or "metadiscourse"). Such referring clearly exists in language, e.g. *He likes Mozart and Bach, but I prefer the latter*, where *latter* refers to the position of the word *Bach*. The possibility for such referring is inherent in the ordering of word entry, mapped (though with various changes, section II) onto the linear order of words in a sentence. The presence of metatextual operators may be seen if we consider, say, *A glass tipped and a glass fell*. If we want to give the information that the two glasses are the same, it is possible to use the inherently available address: *tipped* and *fell* are the first and second arguments, respectively, of *and*; *glass* is the (first) argument of each of these. Then under *and* we address *tipped* as 1, *fell* as 2, the first *glass* is 1.1 and the second is 2.1. The information about sameness here would be given by having on *and* an operator *has the same referent* (or *indicates the same individual*) whose arguments are the contents of the addresses 2.1 and 1.1. With this sameness-operator we obtain reductions to *A glass tipped, and it fell* or to *A glass tipped and fell*. These operators may seem to be peculiarly complex and merely a figment of theory, but they are inescapable for a simple analysis of language. For, as will be seen in section II, such operators yield in a natural way the phenomena of zeroing and pronouncing, without appealing to any grammatical apparatus beyond what exists otherwise in grammar. Without these operators, one would have to announce the fact and conditions of zeroing and pronouncing in various grammar-statements, i.e. statements made in the grammar about the sentences of the language. However, if we consider such grammar-statements (which are themselves English sentences), then we see that their information can be given by metatextual operators of English attached to the very

sentences which the statements are describing. Naturally, the information about sameness cannot be given until the two words which are the same have both entered the sentence: hence we need an O_o operator (*has . . . same . . .*) on the O_{∞} (*and*) which had brought the two words together in the sentence. This O_o sameness-operator can also act on O_{nn} (e.g. on *John washed John*) in which case it leads not to zero but to *self*-pronouns: *John washed himself*. And when the sameness-operator specifies which words are the same, the simplest way of addressing them is by their entry-order in the operator history of the sentence; the word-order in the sentence derives from this, but in a complex way.

Particularly important are the *wh*-pronouns, which create the relative clause, and all those sentence-segments that the grammarians would call modifiers—whether on noun, or verb, or sentence, or whatever. They also underlie all cross-reference. Operators on a sentence can become parts of a “noun-phrase” or any similar “word-phrase” (i.e. a word with its modifiers) only via these. Syntactically, these *wh*-words are the form that pronouns can take when they follow semicolon, which is an O_{∞} connecting two sentences: *The man who had been here left* from semicolon on a sentence pair: $\leftarrow$ *The man left*; *The man had been here*, with pronouncing of the second *man* into *who*.⁷ The pronouncing, of course, arises from a sameness-operator on the two occurrences of *man*: *The man left*; *the man had been here*; 2.1 indicates the same individual as 1.1. Similarly, *The man whom I saw left* is a reduction of *The man left*; *I saw the same-as-1.1 man*. $\leftarrow$ *The man left*; *I saw the man*; 2.2 indicates the same individual as 1.1.

Introducing a sameness-operator as a base for zeroing and pronouncing makes it possible to obtain these as reductions from a metatextual operator inside the base sentence (section III) with no apparatus of operators and reductions beyond what is in any case needed to obtain the referential effect of words such as *latter*. Nevertheless, it makes the base sentences look artificial. Instead of this, one could therefore take pronouns not as metatextual changes (*it* for *glass*), but simply as words of

referential or deictic meaning: *he* meaning “a known or mentioned male person,” etc. And instead of some, but not all, zeroing we could have *and* acting on operators or arguments before these meet (*tipped and fell* acting together on *glass*). Such analyses would make the base closer to the usual language, but would reduce the simplicity of the system that can produce all sentences.

5. Likelihood-Inequalities

The crucial relation in language structure, after the sentence-making ordered entry, is the fact that every operator-word has a particular likelihood-gradation, more precisely inequalities of likelihood, for the words in each of its argument positions, and vice versa. Since we can hardly measure the likelihoods of, say, each *N*'s occurring under *slept*, and so on, we have to deal with the estimate of likelihood as made by speakers of the language, i.e. with the presumed frequency of occurrence. Even so, no precise data can be obtained for all *N* in a particular position, but it is sufficient to work with several grades such as especially low (vanishingly small) likelihood (e.g., under *slept*, for *vacuum*), rather low likelihood (e.g., for *ocean*, *rock* as in *The oceans slept*), somewhat-lower-than-ordinary likelihood (*The flower slept*), ordinary likelihood (*The boy slept*); the last (ordinary) is often called the selection for the given verb in the given position. In addition, some operators have words which are exceptionally likely to occur as their arguments, e.g., *to be here* as argument of *expect* (sect. 7). One may think of intermediate grades, as for *The tree slept* (just below ordinary likelihood?), and one may be uncertain in some cases (e.g., does *earth* have ordinary likelihood under *slept*, as in *The earth slept under the blanket of snow*). The gradation is, of course, related to meaning, but by no means precisely, not only because of meaning extension and explicit metaphors, but also because of the special frequency of certain literary expressions (e.g., *The earth slept*), and so on. While the likelihoods themselves change readily in time and differ as among speakers, the gross classification is more stable. Above all, the inequalities in likelihood are preserved under transformations (below).

⁷ The alternative would be to say that the segment headed by *wh*- is something new, to be called a modifier, which is added directly to a noun: as though *The man who was here left* is formed from *left* operating on *The man who was here*, and *The man whom I saw left* is formed from *left* operating on *The man whom I saw* (these noun-phrases being formed in turn from *who was here*, *whom I saw* being added to *the man*). But if we consider all possible *wh*-modifiers on a particular noun N_1 we find that they are simply all sentences containing N_1 , with N_1 omitted: e.g. *was here*, *I saw*. There is no independent structural characterization of what segments can be added to N_1 ; we have to say that these segments are sentences which contain N_1 and in which the N_1 has received zero (or *-om*, etc.) shape.

Instead of speaking of inequalities of likelihood for the arguments of an operator, one can speak of what further operators on the given operator (which, as in sect. 7, includes what context of neighboring sentences, etc.) would make a particular word not low in likelihood. For not otherwise specified likelihoods, this context is ordinary speech and writing. For specialized word uses, e.g., in science, this context is the neighboring material of scientific articles or conversations. For certain kinds of rare word uses, such as in *The saucer laughed*, the context can be neighboring fairy-tale sentences. And for ungrammatical word sequences (i.e. ones not satisfying the argument-demand relation) the context is a metalinguistic operator such as *word* (as in *He said "Go the a,"* reduced from *He said the words "Go the a"*; note that one cannot say $\exists$ *He said that go the a*, which is not reduced from a zeroed operator *word*, and in which *go the a* would be therefore occurring not under a metalinguistic operator).

Two items of background should perhaps be given about this appeal to likelihood:

(1) Unstructured attempts to characterize the likelihood of word-combinations have failed when they related merely to neighboring words in a sentence, or the like. But when we deal with relative likelihood (i.e. inequalities of likelihood) of arguments in respect to their operator, or vice versa, the likelihood differences begin to fit closely with meaning differences, and to be preserved under sentence transformation.

(2) Likelihood inequalities of operator-argument combinations do an important part of the work of indicating the grammaticality of sentences, i.e. approximately the work of characterizing which sentences are in the language. Students of language have avoided considerations of frequency and likelihood, because these were thought to be unrelated to structure. Thus, there are some sentences which one is certain were never said, but which are immediately understood as grammatical: e.g. *Some blue and mauve onion-skin shot through the air at 759.06 miles per second*. However, such unlikely yet fully grammatical sentences have a particular relation to likely sentences: likely sentences can be formed immediately from these by substituting classifiers for the specific words and adding sentences stating that the specific words are cases of those classifiers. Thus we can form: *Some colored*

solid object shot through the air at a particular velocity; Blue and mauve are colors; Onion-skin is a solid object; 759.06 miles per second is a particular velocity.

There are other kinds of sentences whose likelihood seems unrelated to their grammaticality. Here too, it is possible to find that the different ways and degrees of a given sentence's grammaticality correspond to different sentences of non-low likelihood to which the given sentence is syntactically related. Thus new or far-fetched metaphors, jocular and nonce uses of words, and marginal sentences whose grammaticality is uncertain, can all be characterized as particular kinds of combination of, or departure from, corresponding sentences of ordinary non-low likelihood.⁸

The likelihood-inequalities of arguments for each operator, and vice-versa, differentiate the operator words and their meanings. Some operators have special properties in this regard, by virtue of which they can be collected into subsets in respect to selection. Thus, some have very broad selection, i.e. most or very many of the words in their argument-demand have non-low likelihood of occurring under them: e.g. under *before*, *after* more different words are likely to occur than under *to the right of*, *on top of*. Some have each a favored ("appropriate") argument which is the most likely one under the given operator, e.g. *to be here* under *expect* (sect. 7). For some, what is most likely is not that the argument should contain particular words rather than others, but that there be some word which occurs in each of their arguments, at whatever depth: so for the O_{∞} in sect. 3. There are also cases of various operators having similar likelihood-gradations. Thus the selections of *before* and *after* are very similar though not identical; similarly as between *is more than*, *is less than*, *equals*. Certain O_{∞} (chiefly *and*, *or*) have in common the property of being often semantically commutative and associative; their non-commutative and non-associative occurrences can be obtained from zeroing such words as *then* (sect. 7). That is to say, *He is tired and he is hungry* means much the same as *He is hungry and he is tired*; and if the arguments of *and* are *He is hungry and he is tired*, *He is sleepy* the resultant sentence and meaning are the same as if the arguments are *He is hungry*, *He is tired and he is sleepy*: both yield non-ambiguous *He is hungry and he is tired and he is sleepy*.

There are subsets of operators some of whose

⁸ Zellig Harris, *Notes du cours de syntaxe*, ed. by Maurice Gross (Paris, 1976); and *Transformational system of English: information and reduction*, in Formal Linguistics Series, Reidel, Dordrecht, to appear. These volumes contain data needed to support the present theory as a grammatical analysis.

members have similar selections, while others have contrasting selections. Thus a number of O_0 operators have in common the following properties: they have a broad selection; and the likelihood-gradations of their arguments are much the same under all of them. Other O_0 operators have a likelihood-gradation for arguments which is almost the reverse of that for the set above. Thus under *lasted*, *continued*, we find that *eat* has ordinary likelihood and *arrive* low likelihood: *His eating continued*, but hardly *His arriving continued*; but under *occurred*, *was sudden* the likelihoods are reversed: *His arriving was sudden*, but hardly *His eating was sudden*.

Certain operators, *is a word*, *is a sentence*, *is an O_{no} operator*, etc., are uniquely distinguishable in language analysis, and form the base for the metalanguage, even though they may have in the language itself a status not different from that of other classifiers such as *is a mammal*, *is an atom*. This is because their argument selection is approximately the set of all words, all sentences, all O_{no} words, etc., respectively. However, the arguments are precisely determined only in the field of language-analysis, where we consider only the arguments which together with *is a word* or the like make a scientifically reasonable sentence (*Book is a word*, but not *She spoke is a word*). In the language itself one can find *It is*, *in two words*, *impossible*, and other non-word arguments of *is a word*, just as one can find *A whale is a big fish*. These classifiers, even the metalinguistic ones, do not form clear-cut subsets of operators in the ordinary language.

These various argument-likelihood properties give a texture of similarities to the operators on a given argument-demand. Except as noted at the end of sect. 4, they fall short of creating subsets of operators, because for many of the properties there are operators which have the given property to a lesser degree or in a different manner than the other operators. However, they have the effect of creating semantically important, though fuzzy, categories of the operands: e.g. durative verbs (*eat*) as against momentaneous (*arrive*). And these properties are structurally important in being the basis for the reductions (section II).

It must be stressed that the likelihoods are not used directly in constructing a grammar. The major use is that preserving the inequalities of likelihood—in a specified way—is a proved criterion in determining the transformations of a language (III). For the rest, the discussion of likelihoods has an interpretational status, in characterizing the transformationally-established reductions.

II. REDUCTION

The importance of the sentence-making operations described above—the word-entries satisfying their argument-demands, the argumenthood-indicators, and the likelihood-inequalities—is that all remaining sentences of the language can be obtained from these by a few physical types of reduction in the shape, or relative position (relative distance), of the words which enter a sentence (fn. 8). The great bulk of reductions can be made on an operator, or its argument, if one of these has exceptionally broad selection or exceptionally high likelihood of occurrence in respect to the other, or if it otherwise contributes little or no information at its point of entry in the construction of the sentence. The reductions are thus not arbitrary changes. Most or all have a common property, based on the relative likelihood property of the operator-argument relations, of reducing high-likelihood low-information entries. It is also found that all, or most, of the changes (which will be seen below to constitute sentence transformations) can be best obtained if we assume that the reductions are carried out as the operator enters the sentence or as the specified conditions for the reduction come to be satisfied by the sentence as so far constructed. This greatly simplifies any computational or effective procedure for sentence transformations. It means that if a reduction has taken place on a given operator, the next operator to enter the sentence enters upon the reduced form of the earlier operator.

Not all low-information entries occasion a reduction. These have to be discovered in each language, although some types are widespread. There are three main types of reduction: zeroing and pronouncing for operators or arguments which have certainty or highest likelihood of occurring at a particular point in the construction of the sentence (sects. 6–8); affixation and intonation in the case of operators which have broad selection in the language (sect. 9); moving of metatextual material to the address to which it refers (sect. 10). All reductions, including zeroing, leave a trace which is seen in the entries immediately environing the one which was reduced. Where there is no such trace, no reduction is defined.

6. Pronouncing and Zeroing for Repetition in the Discourse

The metatextual O_0 which assert sameness of two addressed arguments (sect. 4) make one of the

occurrences of the repeated word certain, given the other occurrence, as in the two occurrences of *glass* in sect. 4. The identity of the word at the given location can therefore be dispensed with, as being recoverable from the assertion of sameness on two addresses. In some situations, especially when the other address is unrestricted (i.e. when the other occurrence could be anywhere under the meta-textual operator), the repeated word can receive a shape which is tangible but which is not individual to that word: a pronoun. E.g. for *John's winning was due to John's endless training* (plus sameness-operator) we have both *John's winning was due to his endless training* and *His winning was due to John's endless training*.⁹ In other cases, especially when there is a fixed entry-relation, hence fixed initial distance, between the two occurrences of a word, the repeated word can receive zero shape.¹⁰ E.g. from *A glass tipped and a glass fell* (plus sameness operator) we have *A glass tipped and fell*; from *He will buy a book if she will buy a book* (where the sameness-operator is stated about the entering words, not the individual referents) we have *He will buy a book if she will*. Although the occurrence of the repeated word is certain, given the sameness-addresses, different sentences with different sameness-addresses may yield the same pronoun locations, or the same word sequences after zeroing. Hence ambiguities arise among sentences, i.e. degeneracies of word-sequence in respect to the ordered word entries: e.g. *I left him feeling sad* from *I left him, I feeling sad*, and also from *I left him, he feeling sad*.

7. Zeroing for Highest Sentence-Entry Likelihood

When an operator enters upon its arguments to make a sentence, the argument can be zeroed if it is by far the likeliest (the "appropriate") argument for the given operator; or the operator may be zeroed as most likely or appropriate on its arguments.¹¹

A simple example is the argument of *expect*. The second argument of *expect* is generally an operator

(equivalently, a sentence): *We expect their departure at 6*, *We expect them to depart at 6*. However, we also find certain elementary arguments as second arguments: *We expect John*. The inequalities of likelihood for *N* as second argument of *expect* are approximately those for *N* as argument of *is here* or the like: *John is here*, **Time is here*, **We expect time*. Also, if for each sentence having *N* as second argument of *expect* we form a corresponding sentence having *N to be here* as second argument of *expect*, we find the two corresponding sentences to have approximately the same inequalities of likelihood for various further operators on them, i.e. for various contexts. The two inequality-similarities above are precisely the conditions for one form to be a transform of the other (section III), and we can consider *to be here* to have been zeroed under *expect*, as being the most likely argument of *expect*.

Similarly, under the operator *say*, whose last argument-position can be taken by any operator (i.e. any sentence) but also by metalinguistic classifiers such as *word*, the latter have a uniquely appropriate likelihood and are zeroable. Thus *wh-operating* on the pair *John said words*, *It is late are words* produces *John said the words it is late*, zeroed to *John said "It is late"*; here *words* was the second argument of *said*. But when *It is late* is directly the second operand of *said*, we have *John said it was late*, *John said that it was late*.

A simple example is the zeroing of *then* after *and* if the two arguments of *and* are clearly time-ordered: *He took sick and died* ← *He took sick and then died*. Many complex structures in the grammar are obtained with little or no further ado by the zeroing of especially likely, appropriate, arguments. Examples are the zeroing of *amount*, *degree*, and the like, under quantifier and comparative operators such as *increase*, *is more than*, *is less than*. So also the zeroing of *moment*, *period*, and the like, under time-order operators such as *before*, *after*; this simplifies the grammar of tenses and aspects.

Somewhat differently, an indefinite appropriate

* There is also a special pronoun from the *Speaker says* operator: *I, you* are repetitions of the two *N*, respectively, in the *N₁ says to N₂* which can operate on any discourse or part of it (below): *N₁ says to N₂: N₁ wanted to see N₂ → N₁ says to N₂: I wanted to see you*; with zeroing of *N₁ says to N₂* (section 8, below) we obtain *I wanted to see you*.

¹⁰ This zeroing is almost always on the second occurrence, e.g. in parallel positions under the commutative *O_{oo}* (*and*, *or*). Repetitional zeroing occurs in specific positions of specific argument-demands e.g. also under *O_{oo}* when the first argument of the operator which has become an argument is the same as the first argument of the operator on it: for *John prefers for John to stay in* (under a sameness operator) we have *John prefers to stay in*.

¹¹ In many cases, an operator is zeroed only after a further operator has acted on it. The zeroing of an operator when it is the last operator to act would in many cases leave no trace of its having been present; we might then have no evidence that it had been zeroed. The occurrence of zeroing can also be restricted by demands on the word-sequence. Thus in English, indefinite second arguments are zeroable, but not (in general) indefinite first arguments: *I read from I read things*, but not *3 Eat too much from People eat too much*.

second argument can be zeroed under many operators: *I read* is a reduction from *I read things* or the like.¹²

8. Zeroing for Assured Occurrence or No Information

Certain operators, together with a part but not the whole of their operand, are zeroable when they are the only words that could be occurring in the given entry position. This applies to relative-clause *wh*-words plus *is*, as in *The man here has been waiting* ← *The man who is here has been waiting*; and the same with the indefinite *that*, *anything*, or *group* or the like, preceding such *wh*-words, as in *I never eat her cooking* ← *I never eat anything which is her cooking*. This accounts for the difference between *Mozart and Beethoven wrote operas*, obtained by zeroing from *Mozart wrote operas and Beethoven wrote operas*, and *Gilbert and Sullivan wrote operettas* which must have a different source, namely *The team which was* (or: *contained*) *Gilbert and Sullivan wrote operettas*, from *wh*-operating on *A team wrote operettas*, *The team contained Gilbert and the team contained Sullivan*.

When *will*, *-ed*, are introduced onto the first arguments of *after*, *before*, and period intonation onto the first argument of *and*, *for* (sect. 9), then the original portion of the operators, namely *after*, *before*, *and*, *for* can be zeroed: *He did it before* is reduced to *He did it*. Similarly *He was sleepy. And he was hungry* is reduced to *He was sleepy. He was hungry*.

The *I say*, *I report* (or N_1 *says*, where N_1 is the speaker) which can be assumed on every discourse, can for that reason be zeroed. Many difficult forms are explained by this. One can also assume that discourses carry adjoined sentences or operators which give all the situational and definitional information necessary to the understanding of the discourse.¹³ These are zeroable because they are already known to the hearer, and thus their having been present (as necessary explainers of the sentence) is a certainty; the trace of each such defini-

tion is the presence (in the sentence) of the defined word. Among the many simplifications that this assumption brings to grammar is the fact that deictic pronouns (e.g. in *This is expensive*, with *this* referring to something known only from the non-verbal situation) become under this assumption simply cross-referential (repetitional) pronouns: e.g. we can begin with something like *We are considering an object here. The object is expensive* (under the sameness operator), reduced to *We are considering an object here. This is expensive*, to merely *This is expensive*.

Also zeroable are operators (some with their first arguments) which are informationless in given situations, where they are similar to the performatives discussed in the philosophy of language. It has been shown that *He asks whether it is late* is reducible to *He asks: Is it late?*, and *I requested you that you please go* to *I requested you: Please go!* (sect. 9). The *asks*, *requests* occur with various first arguments and tenses: *You asked*, etc., and are not zeroable. However, *I ask*, *I request* are unique in bringing no information into the sentence, and are zeroable. For, saying *I ask you: Is it late?* constitutes asking the question and thus says the same thing as *Is it late?* by itself. Similarly, *I request you: Please go!* constitutes making the request and is thus the same thing as saying *Please go!* by itself. Much in the grammar of questions and imperatives (and wishes, etc.) is simplified when they are taken as reduced from *ask*, *request*, etc.; and the zeroing of these operators is reserved for the case with the informational properties noted here.

9. Broad Selection in the Language

Certain non-elementary operators, which have a good likelihood on almost all operators of the language as their arguments, can attach affixes or intonations to their arguments; as seen in sect. 8, the original operators can then be zeroed. And certain operators which have a broad selection,

¹² The contexts of *I read*, i.e. the further operators on it, show that it means not all reading activities (as in *I read everything*) but reading whatever it is that one would be likely to read. Some operators do not have this second-argument zeroing, and these seem to words which are not likely to be said with an indefinite appropriate object: e.g. we don't say *I wear*, nor is one likely to say *I wear things*.

¹³ As an indication of this, consider the following: A grammar can be looked upon as a device which decides which sequences of phonemes, or of words, is a sentence or discourse of the language, and which is not. But we can take an arbitrary sequence of phonemes (satisfying the phonemic structure of the language) and add to it metalinguistic operators which say that the first few phonemes are a person's name, the next few phonemes (say, ending in a phoneme that can be a tense suffix), are a specialized biochemical term (verb) meaning to carry out some particular laboratory operation, and the remaining phonemes are the name of some new chemical compound. Then the phoneme sequence is a sentence of English. When phoneme sequences are accepted as sentences of English without such explanations, it is because these explanations are known to the hearer. We can assume that they existed as operators on the sentence, no less than in the case above, but were zeroed because they were known, i.e. their presence as metalinguistic operators on the given sentence was assured.

though not as broad as the set above, are reduced to being affixes on their arguments; these affix forms in most cases have no phonemic similarity to the operator, and must thus be justified as suppletive transforms. In detail:

The time-order words, chiefly, *before*, *after*, can impose *-ed*, *will*, respectively on their first argument, *A*, replacing the operator-indicator *-s* on *A*. When the *-s* has not been replaced, *A* is understood as "present" or "timeless". An efficient explanation of both the temporal and the non-temporal uses of the tenses (including their aspectual properties and such special uses as the past for expressing contrary-to-fact) is obtained by deriving the tenses not from subjective time, but from the time-order of an operator to another operator in the sentence. As a rough example, consider (1) *He will have arrived before your return*. For demonstration, we take the *will* as already present, and account first for the *have -ed*. The source would be semicolon plus *wh-* operating on (2) *He will arrive* and (3) *His arriving is before your return* (where *is* is not present tense, but timeless). This produces *He will arrive, which is before your return*. Here the *which is before* permits a past-tense marker to be added to its host *arrive*; and the *which is* is zeroable (sect. 8), yielding (1). As to the tense on the latest-entering operator, here *arrive*, that results similarly from its time-order to the *I say, I report* which can operate on all sentences or texts, and which is later zeroed. We begin with semicolon plus *wh-* on *I report his arriving* and *His arriving is after* (or: *subsequent to*) *my reporting*, producing *I report his arriving which is after my reporting*. On this there operates a metatextual operator stating that the second *report* refers to the first: *I report his arriving which is after my reporting; 2.2 has the same referent as 1.1*. This permits zeroing of *my reporting*; together with zeroing of *which is*, this yields *I report his arriving after* (or: *subsequently*). Independently, the *which is after* permits a future-tense marker on the host *arrive*. With the coming of the tense marker, the argumenthood-indicator changes from *-ing* to *that*, so that *his arriving after* changes to *that he will arrive after*. When the tense-marker is in, the *after* is zeroable; and *I report that* is zeroable, yielding (2) *He will arrive*.¹⁴

Tortuous as this derivation is, it shows that even the tenses express nothing that cannot be expressed by the operators defined in this theory. The tenses satisfy the conditions for being reductions

(transformations) of *before*, *after*, etc., in particular situations.

The operators *ask*, *request* can impose on the sentence which becomes their operand an intonation of question (permuting tense and first-argument word) and command respectively: *He asks whether it is late*. → *He asks: Is it late?* (sect. 8).

Except as above, discourses end in period intonation. In addition, a number of *O_{oo}*, chiefly *and*, *for*, but also *or*, *but*, *because*, *if*, etc., can impose a period intonation on their first argument: *John refused for it was too late*. → *John refused. For it was too late*. Most reductions can then not occur across the period; but pronouncing and certain other reductions can. This is sufficient to segment the discourses into sentences. Then the zeroing of *and*, *for*, by sect. 8, yields *John refused. It was too late*.

Certain broad-selection operators, chiefly ones having the form of preposition or the negative, which enter before an argument or are permuted to that position, can be reduced (usually suppletively) to being prefixes on their argument: *It is under normal* → *It is subnormal*; *It is a half circle* → *It is a semicircle*; *It is less than finished* → *It is unfinished*.

Certain broad-selection operators, meaning *condition*, *state*, *tendency* and the like, take in English a position after their argument (in many cases by the "compound-noun" permutation of sect. 10) and then reduce to suffixes on their argument: *His childhood was happy*, via a non-used **His child-state was happy*, from *His state of being a child was happy*, from *happy* operating on *His being a child was a state*. The argumenthood-indicator makes this into *The state of his being a child, His state of being a child*, as it makes *His being absent is a trick* into *The trick of his being absent*.

It appears that the affixes of English can be related in this way as suppletive (i.e. phonemically dissimilar) transforms of operators on the words to which the affixes are then attached. In some cases (most prefixes, and some suffixes including the *-hood* above), the affixes are historically not suppletive but reduced shapes of those operators (fn. 8).

10. Permutation to Referred Address

Given that each operator enters into the fixed position after its first argument, most permutations in English are cases of a metatextual operator or second argument, which contains an address (sect. 4), moving to the address to which it refers. Thus,

¹⁴ This analysis provides the same base for tense-consecution and other exceptional uses of tenses as for the ordinary tense on the main verb of a sentence.

a semicolon sentence (especially if it contains a *wh*-pronoun) moves to after its referred address: *My friend returned; I had mentioned him to you* → *My friend—I had mentioned him to you—returned*. Also *My friend returned, whom I had mentioned to you* → *My friend whom I had mentioned to you returned*. When something is zeroed in the second argument of *and*, the residue (i.e. *and* with what is left of its second argument) moves to after the last non-referred material in the first argument: *John saw Mary and John phoned Mary* (under the sameness- O_0) → *John saw and phoned Mary*. After zeroing of *which is*, and *who is*, the residue, if it consists of certain sets of short words such as adjectives, moves to before its referred address: *The pen which is blue writes well* → *The blue pen writes well*. In English, if the residue is of the form of a preposition (chiefly *of*) plus noun (including words made noun-like by their affixes), then the preposition is dropped and the residue moved with compound-noun stress to before its referred address: e.g. this occurs twice in *The machinery is for the building of roads* → *The machinery is for road-building*, and *The machinery for road-building failed to work* → *The road-building-machinery failed to work*. The first compound-noun form here, *road-building* is part of an argumenthood-indicator; the second, *-machinery*, is from *wh*: *machinery which is for someone's road-building*.

There are a number of O_0 operators which appear not after their argument but before their own argument-word and after that word's argument in turn: *He stopped running* (as against *His running stopped*); the auxiliaries, as in *He can run*; the negative in *He is not Greek* (compare *His being Greek is not so*, or the like). In some cases there are grounds for saying that O_0 after its argument carried a referent to the subject of that argument, and so moved to it: *His running stopped due to the subject*, or the like, → *He stopped running* (sect. 3). But in other cases, such as the *not*, such an explanation for the permutation is not available.

11. Reductions Unrelated to Amount of Information

There seem to be some reductions, perhaps only permutations, which have no basis in low informa-

tion. Such is the moving of certain short sentence-segments to before long ones: We have *I saw people near him*, but not **I saw near him people*. However, when long modifiers (i.e. *wh*-residues) are attached to *people* we find both *I saw several uninvited people near him* and *I saw near him several uninvited people*. There are also special permutations for certain adverbial words: *So they say*, *Hardly had they come than he left*.¹⁵

III. THE ENTRY-AND-REDUCTION SYSTEM

The essential fact is that the reductions (II), when applied to the ordered-entry discourses (I), suffice to characterize effectively all the sentences of the language. Since the entries are ordered, and the reductions of a word take place at its entry or at the entry of the immediate operator on it, or are referred to these, we have a decision procedure for constructing and analyzing each sentence. For construction: we have particular partially (but mostly linearly) ordered entries satisfying argument-requirement, and reductions satisfying their specified low-information conditions. For analysis: since zeroing is not a loss of a word but simply a zero shape for a recoverable word, each sentence can be directly transformed and segmented into partially ordered reductions and entries.¹⁶ The degeneracies in some pronouns, zeroings, and other reductions, in which different reductions on different operators yield identical word sequences, cause ambiguities; this only means having more than one analysis for the given word-sequence, although determining the various analyses may be difficult. Aside from the local cases of unordering among certain semicolon entries, the entries can be ordered with their arguments in parentheses. And since every word has stated requirements (one or more) as to its argument-sequence, the entry order of words in a particular sentence can be represented in Polish notation.

From the definitions of the argument-demand sets of words, it follows that every sentence must contain at least one elementary argument *N*, which alone can enter without prior entries; and for

¹⁵ The vocabulary also contains certain variants of word-form, called morphophonemic variants, which affect the shape of a word or affix under particular operators or on particular arguments, and which are not reductions and are unrelated to amount of information: e.g. *knife*, *knives*. The obligatory transformations (which are few, if any) and the morphophonemics are presented in the base sentences as sets of variant forms which certain words take under stated entry conditions.

¹⁶ The partial ordering arises, in the case of reductions, from independent reductions on the same entry. Algorithms capable of analyzing the structure of virtually all sentences of English in the sense of the system presented here have been written. Some have been successfully implemented as computer programs which carry out sentence analysis. Complexities in stating the domain of certain reductions, and degeneracies, make sentence-analysis no simple matter.

every n co-entering elementary arguments it must contain one n -argument elementary operator. Hence it contains at least one elementary sentence. And for every operator or operator-pair of whatever kind, it may contain a non-elementary operator on it.

Every operator makes a sentence. Every non-elementary operator acts on a sentence or sentence-pair (namely, on the sentences made by the operators which are the arguments of that non-elementary operator), and makes a further sentence. Every reduction acts upon a sentence, and results in an altered sentence. Thus all non-elementary operators and all reductions make sentences out of sentences, and are transformations in the set of sentences. The non-elementary operators are a set of transformations on the set of base sentences $\{S\}$ as constructed in section 1. Each non-elementary operator acts on all sentences (since there are few restrictions, but rather inequalities of likelihood, on the sentences in its argument positions), mapping the whole set of sentences $\{S\}$ into $\{S'\}$ (onto a subset having that non-elementary operator as last entry), preserving the inequalities. Each O_∞ maps $\{S\} \times \{S'\}$ into $\{S'\}$. The non-elementary operators act also on reduced sentences, not only on the unreduced ones of section I. But in this case there are restrictions, e.g. there is no *and* on the pair: question, assertion; hence here we have only partial transformations. The reductions are a set of partial transformations on the full set of sentences $\{S'\}$, each mapping a subset of sentences (those containing a particular low-information entry) onto another subset (those containing the reduction), preserving the inequalities. The way inequalities of likelihood are preserved under non-elementary operators allows for a limited amount of exceptions: as an extreme example, the operator *not* changes the relative likelihoods of various kinds of sentence (e.g. of general sentences) but preserves the inequalities of the great bulk of ordinary sentences (*John will leave*, *John will not leave*, both normal, as against *Vacuum will sleep*, *Vacuum will not sleep*). In contrast, the preservation of inequalities of likelihood under reductions is much stronger, leaving room for few word-choice exceptions.

Preserving inequalities of likelihood of operators

in respect to arguments guarantees preserving meaning in each sentence, aside from a reasonably small number of exceptions (idioms and the like). The non-elementary operators preserve the meaning of the sentence on which they act, and add their own meaning in respect to it. The reductions preserve the meaning and add no objective information. They are thus more or less paraphrastic transformations. In addition, any individual sentence may have various non-transformational paraphrases based on synonyms and circumlocutions special to its words in their neighborhoods. These paraphrases do not preserve inequalities, and do not remain paraphrases for other word-choices in the corresponding positions.

It is for this reason that these relations were called transformations in linguistics, and the elementary sentences (each produced by an elementary operator, and containing no sentence as a proper part) were called kernel-sentences because they were the kernel of the natural mapping of the set of sentences onto the set of transformations (as a quotient set of it relative to having the same last operator or the same last reduction).¹⁷ Having the same last operator or reduction is an equivalence relation in the set of sentences, as is also having the same ordered entries.

In particular, the reductions give rise to a partition of the set of sentences into equivalence-classes, in each of which all sentences have the same ordered entries and objective meanings. Since almost all the reductions are optional, each equivalence class (with certain adjustments) contains one reduction-less sentence; any obligatory morphophonemics (fn. 15) is included in the reduction-less sentence. These sentences have a distinguished syntactic form (consisting of word-entry only), and the set of them is closed under the word-entry operation: any word-sequence satisfying it is such a sentence. Hence we may call this set a sublanguage. Since the reductions do not materially alter the information in a sentence, this sublanguage expresses all the objective information of the language.

It remains to consider the structural effects of the reductions. The connection of the reductions to low information brings into the language restrictions

¹⁷ To establish this relation, we have to take the set of sentences, $\{S\}$, as a monoid (with null sentence as identity) with *and* as binary composition in it. For any two sentences A, B , we have A and B as a new sentence C . The types of sentence-pairs C, D on which *and* is grammatically unable to operate (e.g. *Are you going?* and *I'm late*), are sufficiently few so that we can reasonably put their resultant as the null sentence: C and $D = \text{null}$. It follows that there is a binary composition in the set of equivalence classes $\{E\}$, with E_A and $E_B = E_{A \text{ and } B}$ (where E_X is the equivalence class to which X belongs). The natural mapping is then a homomorphism of $\{S\}$ onto $\{E\}$.

and subclasses such as had not arisen in the word-entry process. As an example, consider the reductions that produce tense (sects. 8 and IV):

Once the tenses are understood as reduced from time-order words, we can see how the operators which receive the tense affix become specialized into verbs and adjectives. Operators which had a high likelihood of occurring under *before*, *after* receive the affix directly on them, and become verbs: *He will leave after eating*, *He phoned before arriving*. Operators referring to more stable events are less likely to be time-ordered to other operators in the sentence (except to *I say*); these receive the affix indirectly, on a carrier *be*, and become adjectives: *The river is long*, *He was peculiar*. (The *was* here is from *before my saying this*.) While adjectives such as *peculiar* can occur with a *before* which relates them to some other operator in the sentence, as in *He was peculiar before she met him*, they are much less likely to do so than are the operators which become verbs, such as *phoned*.

We can now see why assigning reduced forms to high likelihood, low-information, operators creates restrictions and subclasses. The reason is that operators are complexly graded as to likelihood, whereas a reduction permits in general only two grades: receiving the reduction, and not. Hence a conventional cut-off point must be imposed on the graded operators, as a boundary for recipients of the reduction. As a result, restrictions are created: the reduction is applied restrictedly only to one member of the operator or argument class, or to a subdomain of it whose members have to be listed or characterized. Thus, subclasses are formed: a given set of operators is divided into those that receive the reduction and those that do not. For example, the directness or non-directness of tense attachment was seen to depend upon the operator's likelihood of occurring under *before*, *after*; and it serves to separate out verbs from all other operators. But while the likelihood of an operator's being under time-ordering is graded and in part uncertain, the recipients of direct attachment are a single subset. The inevitable borderline cases are decided somewhat arbitrarily but definitely; *sleep* is a verb, *ill* is an adjective. Other languages have much the same verb-adjective system, but the borderline decisions

may be different: *ill* a verb, *sleep* an adjective.¹⁸

We have here the characteristic properties of social conventions: a use-oriented graded human activity; and an organizing of some feature of it, which in so doing makes arbitrary decisions that have to be maintained by convention. More than this, many of the language conventions, even though no social interests or class control lie behind them, are institutionalized into rules, such that a departure from them is an error. It is here that rules come into grammar: to say *He illed* is a "mistake" in English, not just a nonsensical yet possible sentence such as *Vacuum ate cassettes*.

IV. RELATION TO ISSUES IN PHILOSOPHY

12. Symbols and Terms

The methodological approach which led to the theory of language here consisted in comparing the occurrences of segments of speech and writing—the tokens—relative to each other; it is thus close in spirit to the inscriptional approach. When the occurrence of words in the source, i.e. unreduced, discourses is found to be determined by an entry order satisfying argument-demand, we obtain a classification of words by their argument demand. This is a special case of the idea of functors in categorical grammar. However, it is this particular case that is adequate for language; not every way of defining functors would be equally suitable.

In the various operator classes defined by their argument-demands, particular kinds of terms of philosophical relevance are produced by particular low-information reductions. Thus, most abstract terms in language are nominalizations of operators (i.e. operators with argumenthood-indicators on them) whose own arguments are indefinite nouns which have been zeroed: e.g. *humility* from *the humility of people from the property of people's being humble*.

Disposition terms are, in language, operators under certain further aspectual (modal) operators. Suffixes, including *-ible* and its synonyms, are reduced from operators on the word to which the suffix is affixed: *X is soluble* from *X's dissolving is possible*; and suffixless disposition terms are found to result from a zero reduction of *possible* or the

¹⁸ In some cases, the borderline for a reduction is not so fixed conventionally. This becomes the ground for productivity and for marginal sentences (sentences whose acceptability is uncertain), as in such forms as *The baby took a crawl over to me*, extended from *He took a walk*, etc. Of course, all this information about which words are verbs and which are adjectives, or which verbs can be objects of *took*, would have to be given in any grammar. In the present system, these facts are relevantly organized, and some can be deduced from other facts.

like.¹⁹ In addition, since all non-indicative forms are derived by the theory above from indicative ones, it follows that the subjunctive conditional, which is semantically related to disposition terms, can be derived from conjoined indicative assertions. Furthermore, as among various occurrences of conjoined assertions, it has been found that for a conjoined assertion to have good likelihood of occurrence, or for it to express causality and necessity rather than accidentality of their connection, it is preferred that the same words repeat in the two conjoined statements, or in a chain of intervening conjoined statements which expressed well-known facts and which had therefore been zeroed as contributing little information.²⁰ Taken together, all these grammatical results suggest a syntactic characterization both of disposition terms and of the difference between laws and accidental correlations (as in counterfactuals). The fact that among disposition terms and among subjunctive conditionals there is a gradation²¹ is not surprising, once we see that their linguistic characterization rests on the degree of word-repetition in the chain of intervening sentences connecting the two statements.

A subset of words can be characterized not only by the kind of operators involved, but also by the kind of reduction which the words receive, occasioned by the particular kind of informationlessness the words have. Thus, the unique way in which *I ask you, I say to you*, etc., carry no information in *I ask you: Did he leave?*, etc., permits their zeroing to *Did he leave?*, etc., and thus isolates on syntactic grounds certain types of performatives.

13. Structure and Translation

The investigation of entry order as specifying, before reduction, the structure of discourses and sentences shows that discourses and sentences are not just concatenations of words constrained by one or another set of rules, but a particular sequence of operators on arguments. The meaning of a sentence is closely related to its ordered entries; therefore, translation, even under the conditions discussed by Quine in *Word and Object* and elsewhere, is not entirely indeterminate. For the most part, operators of a given argument structure in one language translate into operators of the same

argument structure in another (verbs on two nouns into verbs on two nouns, like *drink* and *boire*; verbs on a noun and a verb into verbs on a noun and a verb, like *believe* and *glauben*, etc.): there are a few important exceptions, which may be transformational, such as the possibility of translating (even in the same language) O_{nnn} (e.g. *put* in *He put the book on the table*) by O_{oo} on O_{nn} , O_{nn} (e.g. *so that on place, be*, in *He placed the book so that it was on the table*). It is true that the selectional (hence, meaning) range of each word within a class may differ widely in the two languages; and idioms and allusions differ, and there may even be circumlocutions which express a meaning of a particular sentence by a whole sequence of operators of different argument-demands from those in the given sentence. But the way in which the possible meanings of a sentence are constrained by its operator structure limits the indeterminacy of translation.

14. Information vs Truth

The predicate-operator structure of language, and such specific facts as that all non-indicative sentences (e.g. the question) are reduced from indicatives, shows that language is a structure for indicating (indeed, for transmitting) information; it does not have any basic equipment for expressing attitude and emotion, or for distinguishing truth.²² Also, when a further operator enters on a sentence, the resultant retains the meaning-contribution of the original sentence; but there is no necessary relation between the truth value of the original sentence and that of the resultant: *Zurich has a subway, He thinks Zurich has a subway, He denies that Zurich has a subway, Zurich may have a subway*. Not a few of the difficulties in the philosophy of language and in neighboring areas of philosophy arise from starting with the equipment which had been developed for truth systems, and using it to analyze the information system that language presents. Furthermore, where the set-theoretic equipment of logic cannot reach, the custom has been to use subtle but uncontrolled and unsystematic appeals to meaning. More adequate methods are now available.

For example the two alternative grounds for identifying meaning have always been the intension and the extension of words. However, the inscrip-

¹⁹ Note the relation of dispositional predicates to possible occurrences (grammatically: to conjoined statements about these possible occurrences) in Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, 1955), p. 45.

²⁰ See fn. 6.

²¹ W. V. O. Quine, *Word and Object* (New York, 1960), p. 225.

²² *Op. cit.* in fn. 6 (ch. 2).

tional analysis of language reveals another basis (or correlate) for the meaning of a word: its particular selection as to which other words have good likelihood of occurring in operator or argument relation to it. This yields useful characterizations of synonymy and ambiguity, and provides criteria for metaphor and kindred semantic problems. Thus, whereas an extensional view of meaning leaves us with the problem of terms whose extension is null, the problem does not arise in language as analyzed here: *unicorn* and *centaur* have different selections (as, in a novel, do the names of the various characters), no less than *horse* and *dog*.²³ Sentences of good likelihood of occurrence would be, for example, *The unicorn is a mythical animal with one horn*, (1) *The unicorn is graceful*, *The unicorn is depicted in the Bayeux tapestries*, but not **The unicorn is half horse, half man*. Also we would find readily *The centaur is half horse, half man*, (2) *The centaur is a powerful fighter*, but not **The centaur is a mythical animal with one horn*. The asterisked sentences above can be said grammatically; but their likelihood of occurrence is very low, not because they are less true than the others (the truth of (1), (2), for example, is a problem at best), but because that is not what is likely to be said about unicorns and centaurs. One might argue that "likely to be said" is a weak basis for philosophical discrimination. But, given the datum of estimated likelihood of occurrence, the inequalities in it are stable and important entities of language structure. As words, *unicorn* and *centaur* have different selections. And while unicorns do not exist as objects, *unicorn* exists as a word. For words which have non-null extension, e.g. *horse*, the selection of the word is closely related to the properties of the referred object, horse. Similarly, the unique selection of the word *unicorn* can be semantically interpreted, without determining the truth of the properties of the object, unicorn.

Appeal to selection is not merely a device for the resolution of a particular problem. And it is not surprising that the semantic correlate in language should differ from that in set-theoretic systems where the empty set is unique in respect to what can be predicated of it. Speaking about unicorns can have meaning, and differently than speaking about horses—or about centaurs.

15. Quantifiers, Variables, Reference

The most glaring differences between the sen-

tential forms of logic and those of language lie in the use of variables, and of quantifiers on them. Quantifiers are not really a machinery *sui generis*. They merely use cross-reference among variables in a way that gives over-riding status to a particular predicate—namely whether the set is empty or not, or is completely covered by cross-referenced predicates; this meets the needs of material implication. In so doing, quantifiers fill a syntactic role comparable in importance to that filled in many languages by tense, which gives over-riding status to another particular predicate—namely, the time-order between conjoined predicates. The meaning that quantifiers contribute is in some cases simply not expressed in language (e.g. there is no difference between *Opossums have porches* and *Unicorns have horns*); in other cases it is expressed by aspectual (modal) operators on the predicate (as in *Cats sometimes have small tails*, equivalent in meaning to *Some cats have small tails*); and for the rest it is expressed by modifiers on the argument (as in *some cats*); this last uses the same grammatical technique, of cross-reference among arguments, as is used in quantifiers.

More generally, cross reference among the arguments of predicates is achieved in logic by the use of variables within an explicitly or implicitly stated scope; in language, by the use of pronouns. The syntactic basis and resultant properties of these pronouns will be seen in sect. 17, from which it will be seen that the techniques of logic and those of language are not essentially different in this respect.

Reference other than cross-reference is not syntactic but semantic, and is not relevant to the present discussion.

16. Metalanguage and Indirect Discourse

The other great difference between language and logic, aside from that relating to information as against truth, is the fact that the metalanguage is a proper part of the language, more precisely that metalinguistic sentences are themselves sentences of the language. Clearly, *Sentences are sequences of words* is a sentence of English. Furthermore, metalinguistic operators can operate on the language-material to which they refer: "*Water*" is an English word (which linguistically is only *Water is an English word*, such quotes not being phonemic). Because the operator-selection of the metalinguistic operator is very different from that of its argu-

²³ Cf. Nelson Goodman, "On Likeness of Meaning," *Analysis*, vol. 10 (1949), pp. 1-7; Israel Scheffler, "Ambiguity, An Inscriptional Approach" in *Logic and Art*, ed. by Richard Rudner and I. Scheffler (Indianapolis, 1972), pp. 251-272.

ment, the presence of the metalinguistic operator is usually obvious from the further operators, e.g.: *Water has many impurities* as against *The word water has five letters*. Hence the metalinguistic operator is zeroable, yielding (phonemically) *Water has five letters*. The syntactic confusion of use and mention in language is thus due to the zeroability of the metalinguistic operator.

Operators of a metalinguistic character are also zeroable in other conditions, where they lead to the well-known problems of indirect discourse. First, *I ask you whether Mary will leave* is reducible to *I ask you: Will Mary leave?*, where *I ask you* is zeroable (sect. 8 and 12) to *Will Mary leave?* Similarly *I say to you that Mary will leave* is reducible to *I say to you: Mary will leave*, and then to *Mary will leave*. Other segments, such as *I said to you, He said to her*, are not zeroable, but the reduction from *that* to quote-intonation (as separate sentence) is available: (1) *He said that Cicero denounced Catiline* → *He said: Cicero denounced Catiline* (to use Quine's example). Now, *I say* is zeroable also in certain positions which happen to be permuted into the operand of some *X says*. Thus if we have (2) a sentence produced by the *wh-* (relative clause) operator on the sentence pair *I say that he said that someone denounced Catiline* and *I say that someone is Cicero*, then both occurrences of *I say* are zeroable, and we obtain (3) *He said that someone who is Cicero denounced Catiline*, where *someone who is* is zeroable,²⁴ yielding (4) *He said that Cicero denounced Catiline*. Had the second *I say* not been zeroed, the sentence (2) would be *He said that someone who I say is Cicero denounced Catiline* instead of becoming (3). Of course, if the speaker himself had used Cicero's name, the sentence would have been (1). If the speaker did not know that the person he named was Cicero, (1) is ambiguous with (4): the confusion is in whether he said *Cicero* or *I say Cicero*, and the confusion arises from the zeroability of *I say*.

These are straightforward and common transformations in English. What has been done here was to separate the indirect discourse problem into two parts: one, the intonational change between a direct quotation and the identically worded indirect discourse—certainly always possible; two, the changes in the indirect discourse which are due to a *which I say is X* with zeroing of *I say* (and of *which is*). This analysis can be extended to any paraphrase, introduced by *I say*, of the original speaker's words. And it can be extended to cer-

tain verbs other than *He said* (e.g. *He believes*, even *He saw*). Under other verbs there is no confusion: in (5) *I say that he tore the page* there is no different meaning as against *I say that he tore what I say is the page* which would reduce to the same word-sequence as (5).

17. Metatextual Operators and Cross-Reference

As has been noted above, sentences can carry operators which refer to locations within the sentence. This is so because every sentence has a linear order of words and, more important, a partial order of entry, and every sentence is complete before the next higher operator acts on it: e.g. *A man entered* and *A man then phoned* are complete before and operates on them to form *A man entered and a man then phoned*. An operator can therefore give information about ordered entries in its completed operand. By far the most important case of this arises when the operator gives information about sameness of word or sameness of referent at two entries in its operands: e.g. when *1.1 has same referent as 2.1* operates on *A man entered and a man then phoned*, to produce *A man entered and a man then phoned*; *1.1 has the same referent as 2.1*. Here, permutations and zeroings yield *A man entered and the same (as 1.1) man then phoned*, which reduces (preserving meaning, of course) by pronouncing to *A man entered and he then phoned*, and by zeroing to *A man entered and then phoned*. Language thus has a metatextual machinery for cross-reference, for stating that a word in one argument (of an operator) is the same, or has the same referent, as a word in another argument (of that operator). This method rests essentially upon containing the material in question within the arguments of the operator—what one may call citing the material. Otherwise, the metatextual operator would not be able to name the locations of the material which is same. These locations are named as addresses relative to the metatextual operator; no other way is available for identifying locations. This citing is comparable to the scope that has to be stated for the cross-reference between variables in logical formulas, and is the basis of the comparable restrictions on cross-reference (including that inherent in quantifiers) in logic and in language—e.g. that it can be carried out only between arguments, or operators that have become arguments under other operators.

As an example of particular interest, we consider here the characterization of impredicative sen-

²⁴ *Op. cit.* in fn. 8.

tences: Aside from the light which the cross-reference system of language throws upon the syntactic machinery of cross-reference in logic, the particular conditions which make cross-reference (pronouncing) possible in language impose certain limitations upon what can be pronounced. Since nothing can be pronounced which has not been cited, as a complete linguistic entity, under the given metatextual operator, it follows that the self-referring pronoun of impredicative sentences does not exist in natural language. For example, in *This sentence is false*, the *this* cannot refer in English to the sentence of which it is part. The would-be impredicative sentence above is syntactically analyzed as *wh-* operating on *A sentence is false*; *A sentence is this (one)*; this reduces to *A sentence which is this is false*, and then to *This sentence is false*. If we now seek the antecedent of the *this*, i.e. what the *this* refers to, we find that the occurrence of *this* here has to be a reduced form of a repetition of some cited sentence. We would have (disregarding certain details) as source: S_1 . *A sentence which is S_1 is false*. This is pronounced to: S_1 . *A sentence which is this (one) is false*; then permuted to: S_1 . *This sentence is false*.

One could try to avoid this anaphoric analysis by saying that *this* is deictic in *This sentence is false*. But in the theory presented above, deictic pronouns are derived from cross-reference pronouns, in the special case when one of the cross-referenced locations is in a certain kind of zeroable metalinguistic sentence. Consider *This weather is just right*. We begin with *wh-* on *I assert that the weather is just right* and some metalinguistic sentence such as *The weather is the topic of my assertion*, producing *I assert that the weather, which is the topic of my assertion, is just right*. On this we have a metatextual operator, roughly 2.2 has the same referent as 1.1. This produces roughly

**I assert that the weather, which is the topic of my assertion here, is just right*, where the topic of my assertion here has a reduced form *this*. Hence **I assert that the weather which is this is just right* → *This weather is just right*. Such derivations may seem unreal for language, but they use no more than the otherwise established equipment of grammar; and they show that even such special entities as the deictic pronouns can be derived from the simple operator system proposed here.

When the derivation above is applied to semicolon plus *wh-* on *I assert that a sentence is false* and *A sentence is the topic of my assertion*, we obtain *I assert that a sentence which is the topic of my assertion here is false*, reducible to *This sentence is false*. Here, if the sentence in question is false then my assertion is true, and if the sentence is true my assertion is false; no antinomy arises in this derivation in natural language.

The comments above relate only to the cross-reference syntactic derivation of anaphoric and deictic *this*, and to the absence of self-referring *this* in language. However, extensions of this argument may apply to the antinomy of the Liar in general. In the *Wahrheitsbegriff*, Tarski refers to Lukasiewicz' formulation of that antinomy in terms of an empirical statement (page and line on which the sentence is printed).²⁵ Henry Hiż has argued that this empirical statement is based on a deictic pronoun, even if the latter is not explicitly evident. The extension of the argument above to other forms than the explicit impredicative requires further details on cross-referencing in language.

In the same vein as above, the view of language presented here suggests other syntactic investigations that may be of interest to language-related issues in philosophy.

University of Pennsylvania

Received September 26, 1975

²⁵ Alfred Tarski, *Logic, Semantics, Metamathematics*, tr. by J. H. Woodger, *The Concept of Truth in Formalized Languages* (Oxford, 1956), p. 158.

II. SELF-WARRANT: A NEGLECTED FORM OF PRIVILEGED ACCESS

WILLIAM P. ALSTON

Part I

SELF-WARRANT AS A MODE OF PRIVILEGED ACCESS

IT seems obvious (to most of us) that in some way a person has epistemically privileged access to his current sensations, imagings, feelings, and thoughts, to his current *conscious states* (C's). But in just what way? In seeking to answer that question we must first make the concept of privileged access more explicit.

In the most general terms, to say that someone has *epistemically* privileged access to some subject matter is to say that he is in an unusually favorable position to acquire knowledge concerning that subject matter. The subject matter under discussion here is the state of consciousness of a person at a given time, and the privileged access in question is not just an unusually but a *uniquely* favorable position. Thus we may take the following as an unspecific formulation of the Privileged Access Thesis.

- (1) Each person is in a uniquely favorable position to acquire knowledge of his own current conscious states.

The unspecific element in (1) is the term "favorable position." Just what is favorable about this position? On this issue, as on so many others, philosophers have taken every logically possible position, and perhaps some that are not. (a) The most common view has been that one's unique position consists in an immunity from one or another of the conditions we are concerned to avoid in our quest for knowledge. Those most often alleged are immunity from *ignorance* (omniscience), from *error* (infallibility), from *doubtfulness* (indubitability), and from *refutation* (in corrigibility). (b) Others hold that although one does not enjoy any of these immunities one does have *unique resources* for acquiring knowledge of his own current C's, e.g., a faculty of introspection or "inner percep-

tion," that gives him and only him a *direct* awareness of his own C's. (c) Still others deny that there is any sort of uniquely privileged access. (d) And finally there are those who dismiss the whole problem, on the grounds that it makes no sense to speak of a person *knowing* what he is currently thinking or feeling.

I shall be taking a position that falls somewhere in the middle of this array, thereby exposing myself to fire from all sides. As against (c) and (d) I accept the privileged access thesis. As against (a) I take it that we do not enjoy immunities as strong as those commonly alleged. But as against (b) I suggest that the alternative to those excessive claims is not to be found by making explicit the way in which I *come into possession* of knowledge of my own C's. In this paper I shall not be defending my rejection of those positions. Instead I shall be presenting and defending my own version of privileged access, *self-warrant*. Self-warrant involves a kind of immunity, immunity from lack of warrant (justification),¹ that is weaker than the more familiar kinds listed above, while at the same time constituting a respect in which one is *in principle and necessarily* in a very strong epistemic position.

Self-warrant attaches to beliefs. To call a belief self-warranted is to say that it warrants (justifies) itself; more soberly:

- (2) Belief *b* is self-warranted = *a*, *b* is warranted just by virtue of being *b* (being *b* is sufficient for *b*'s being warranted).

Thus a self-warranted belief enjoys an immunity from lack of justification; it cannot be the belief it is and fail to be justified.

We can give a more illuminating formulation by making it explicit that a belief will be self-warranted, when it is, by virtue of some of its features rather than others. If my belief that I am now feeling tired is self-warranted, it is not

¹ These terms will be used interchangeably.

by virtue of its being a belief about a psychological state or even about a *C* (not all such beliefs are self-warranted), but by virtue of its being a belief of a certain person that *he* is *currently in a certain C*. (Call such beliefs *B*'s.) Indeed, this point is not restricted to *self-warrant*. Whenever a belief is epistemically justified in any way it is by virtue of some of its features rather than others. If my belief that a burglar entered my house during my absence is justified, it is by virtue of its being supported by adequate evidence rather than, e.g., by virtue of its being a belief about my house or its being strongly held. When a belief is warranted it is always by its possession of some *general* feature, one shareable by other beliefs. Since epistemic warrant is a position on a dimension of evaluation, we can express this point by saying that the evaluation of a belief as warranted always presupposes some general evaluative principle to the effect that all beliefs with a certain property are warranted.

To return to self-warrant, one can embody the foregoing in the following revised definition.

- (3) Belief *b* is self-warranted =_{df} *b* belongs to a type of beliefs, *K*, such that any *K* is warranted just by virtue of being a *K*.

Since this paper will be explicitly concerned with a certain *type* of beliefs, we may as well define the term for that application.

- (4) Beliefs of type *K* are self-warranted =_{df} Any belief of type *K* is warranted just by virtue of being a belief of type *K*.

To avoid triviality we must observe certain restrictions in our choice of types. Consider the type—beliefs for which the believer possesses adequate evidence. Clearly any belief of this type is justified just by virtue of being of that type. But we should not want to apply the term “self-warranted” to a belief that is justified in this way. (Similarly the fact that every true belief belongs to a class—true beliefs—such that it is impossible that a member of that class be mistaken, does not imply that every true belief is infallible.) I will not attempt in this paper to formulate criteria that will exclude all unwanted belief-types. Clearly the claim that *B*'s are self-warranted is not trivial, and the same can be said for other interesting candidates, e.g., simple arithmetical beliefs.

As just suggested, beliefs of various types

might conceivably be self-warranted, but in this paper we are concerned to defend the claim that:

- (5) *B*'s are self-warranted.

This is the positive part of our version of (1). It says that I am in a favorable epistemic position vis-a-vis my own current *C*'s in that whenever I believe that I am currently consciously thinking or feeling so-and-so, that belief is *ipso facto* justified and hence runs no risk of not being justified. But we still need to add the negative clause—that no one else is so situated. We can't do that by adding “and no one else's *B*'s are self-warranted”; that would contradict (5). It is not in this that my privilege resides. The point rather is that no one else's belief that has the same propositional object as one of my *B*'s is self-warranted. Hence the complete privileged access thesis (self-warrant version) will read:

- (6) For any subject, *S*, every *B* of his is self-warranted, and no belief by any one else with the same propositional object as that *B* is self-warranted.

In taking (5) to spell out a way in which one is in a uniquely favorable position to get *knowledge* of one's current *C*'s, I am supposing that justifiably believing that *p* goes at least part way toward knowing that *p*. If we add the uncontroversial requirement of truth, this means that I am supposing the following familiar conditions to form at least part of a sufficient condition for *S*'s knowing that *p*.

(Belief) *S* believes that *p*.

(Justification) *S* is justified (warranted) in believing that *p*.

(Truth) It is true (it is the case) that *p*.

However I am not committed to accepting the True-Justified-Belief (TJB) *conception* of knowledge, according to which *S*'s knowing that *p* can be defined by the conjunction of these three conditions. Nor am I committed to the claim that these conditions are sufficient by themselves.

Even the more modest claim has been rejected by some epistemologists. Some have maintained that knowledge has nothing to do with belief, and others have denied that it has to do with justification. I believe that an analogue of the concept of self-warrant could be formulated for at least some of these alternative views. but I will not be able to go into that here. Note that

one who denies that justification has anything to do with *knowledge* can still consider whether *B*'s are self-warranted, provided he countenances the concept of epistemic justification at all.

Part II

PRELIMINARY ELUCIDATIONS

Before considering the reasons for and against (6) I shall make a few comments designed to clarify the notion of self-warrant and to set it in a wider context.

A. The Concept of Belief

Some readers may feel that, whatever may be the case with other knowledge, one's knowledge of his own *C*'s does not involve belief. This view may be held in a more or less radical form, on each of which I shall briefly comment.

First it may be held that it makes no sense to speak of a person believing that he is aware of a yellow patch or that he is thinking about the election. But even if it is nonsense to say this in the common sense of "believe" in which it stands in contrast to "know" ("You *believe* your car is out front! Can't you *see* whether it is or not?"), I am not using "believe" in any such sense. Without attempting a full dress analysis, I can indicate the relevant sense by saying that a sufficient condition of *S*'s believing that *p*, is that *S* be prepared to assert that *p* if the question is raised as to whether *p* is the case and if he is disposed to give a sincere and candid answer.² This condition would normally be satisfied by a person with respect to propositions attributing his current *C*'s to him.

Second, granting the existence of such beliefs, one may still feel that one's normal knowledge of one's *C*'s has nothing to do with belief, but simply consists in the state's being "given" or "presented" to his awareness. In Part III, C. 2, I suggest that *C*'s are *not* "presented" to one's awareness. But, even if they are *S* will *also* justifiably believe, in the sense just indicated, that he has the *C* in question. So whether or not there is some non-propositional sense of "know"

in which his being presented with the *C* constitutes his knowing it, he will also have a *true justified belief* that he has the *C*, and it is with the latter that we are concerned in this essay. Residual resistance at this point may come from the supposition that every belief is consciously entertained for as long as it is possessed. But no such condition is required for a person to have the disposition to assert that *p*, in terms of which we explained our sense of "belief." In this sense each of us at any moment has many, perhaps indefinitely many, beliefs, only a tiny fraction of which "penetrate" into consciousness. Thus in saying that a man not only is aware of a yellow patch but also believes that he is aware of a yellow patch, we are not imputing any extra phenomenological complexity to his current condition.

B. Self-Warrant and Other Epistemic Immunities

In rejecting omniscience, infallibility, indubitability and incorrigibility, while accepting self-warrant, I was presupposing that the latter does not entail any of the former. I shall now do something to support that presupposition.³

Omniscience belongs to a different family of concepts from the other immunities. It has to do with what is possible or impossible, given the truth of *p*, rather than with what is possible or impossible given one's belief that *p*. More specifically, to say that *S* is *omniscient* with respect to a range of propositions, *L*, is to say that for any proposition, *p*, that belongs to *L*, it is impossible for *p* to be true and *S* not know that *p*. There should be no temptation to suppose that self-warrant entails omniscience. From the premise that *B*'s are self-warranted we cannot infer the impossibility of a *C* occurring without its possessor knowing that it is occurring. The latter would obtain if the subject did not form a belief that the *C* was occurring; and even if that should happen it could still be true that whenever such a belief is acquired it is *ipso facto* warranted.

The other immunities are like self-warrant in attaching to belief; they all have to do with what is rendered impossible, given that a belief of a certain type is acquired. Within this group the *TJB* conditions for knowledge provides a

² I have termed this *only* a sufficient condition because of the uncertainties of its application to believers without linguistic capacities, like infants and dogs.

³ A much more extensive discussion of these immunities is found in my article, "Varieties of Privileged Access," *American Philosophical Quarterly*, vol. 8 (1971), pp. 225-251. However the points made in this section differ somewhat from anything in that essay.

simple scheme for relating infallibility and self-warrant.

Infallibility—Satisfaction of the belief condition and failure of the truth condition is not a possibility.

Self-warrant—Satisfaction of the belief condition and failure of the justification condition is not a possibility.

This makes it clear that self-warrant does not entail infallibility. It is fundamental to the concept of epistemic justification that a belief may be justified without being true. Hence from the premise that a belief cannot fail to be justified we cannot infer that it cannot fail to be true.

Indubitability and incorrigibility do not fit so neatly into the scheme, since they have to do not with what conditions must be satisfied given the belief condition, but rather with the epistemic positions that are possible vis-a-vis the truth condition, given the belief condition. Indubitability of a belief (as usually understood) involves the impossibility of *grounds for doubt* that the belief is true, while incorrigibility of a belief consists in the impossibility of anyone's showing that the belief is false. Now if something counts as a ground for doubt that p only if it is a ground (perhaps very weak) for believing that not- p , we can discern a simple whole-part relation between what is declared impossible by incorrigibility and by indubitability. If the belief that p is indubitable then there can be no grounds, however weak, for not- p ; but if the belief is incorrigible and not indubitable, something less inclusive is ruled out, viz., strong enough grounds for accepting not- p . Since what is ruled out by incorrigibility is included, as a limiting case, in what is ruled out by indubitability, if we can show that self-warrant does not rule out the former, we will also have shown that it does not rule out the latter. In other words by showing that self-warrant does not entail incorrigibility we will also have shown that it does not entail indubitability. So let's concentrate on incorrigibility.

The argument again rests on very general epistemic considerations. It is a familiar fact that there can be adequate justification for both sides of a contradiction. Examples are legion. I can be amply justified in perceptually judging that there is a desk in the next room, while someone who knows about the unusual arrangement of mirrors is even more justified in believing that there is no desk there. The detective has adequate reason for believing that the butler did it, until additional evidence turns up. If the self-warrant of a belief that p rules out the possibility of someone's having adequate warrant for not- p , it must be because of special features of *self-warrant*, over and above the features of warrant in general. I can see no features that yield this conclusion; and so I conclude that self-warrant is logically compatible with being shown mistaken.⁴

C. Self-Warrant as a Form of Immediate Justification

Self-warrant is a particular form of immediate (direct) justification. We can throw further light on the concept by exploring its relation to other forms.

Where what justifies S 's belief that p includes⁵ S 's possession of some other knowledge or justified belief that is appropriately related to p (S 's having "adequate evidence" or "sufficient reason" for p), we may say that S 's belief that p is *mediately* (indirectly) justified, and, if the other conditions for knowledge are satisfied, that S *mediately* (indirectly) knows that p . Whereas if what justifies S 's belief that p does not include anything of that sort, we may say that S 's belief that p is *immediately* (directly) justified, and, if the other conditions are satisfied, that S *immediately* (directly) knows that p . On this approach the fundamental distinction is between mediate and immediate justification, the distinction between mediate and immediate knowledge being derivative from that.

In view of persistent confusion between a belief *being* justified and someone's *showing* a

⁴ Of course it may be argued that B 's cannot be corrected because no one else can have adequate grounds for a belief about my current C 's, except by taking my word for it. I do not accept this argument, but in any event it is irrelevant to the present issue, which has to do with what is entailed by self-warrant as such, regardless of the content of the self-warranted beliefs.

⁵ "Includes" rather than "is" since various other requirements are often imposed, e.g., that S 's belief that p be "based on" this other knowledge, and that S realize that the latter constitutes adequate grounds. Whether or not these additional requirements are adopted it is clear that mediate justification is distinguished by the requirement mentioned.

belief to be justified by *specifying* what it is that justifies it, let me say that it is the former that is under consideration here. As indicated earlier, I am supposing that for a given type of belief, e.g., singular perceptual beliefs or empirical generalizations, there are principles that lay down conditions under which a belief of that type is justified. When I speak of "what justifies a belief" I am thinking of what will satisfy such a requirement. It is then a further question whether anyone notes, or shows, that the requirement has been satisfied. Sometimes the principle in question requires, as a condition for justification, that the believer have certain other knowledge or justified beliefs; sometimes not. In the former case satisfaction of the requirement will constitute mediate justification; in the latter case, immediate justification.

We should not confuse what is required for justifying a belief with what is required for possessing it. Thus it has been maintained that one cannot have the concept of red as a property of physical objects, and hence cannot believe that a ball is red, unless he knows in what conditions things that look red *are* red. But even if this is so it would not prevent a perceptual belief that a certain ball is red from being directly justified. For the distinction between mediate and immediate justification has to do with what is required for a given belief's being justified *rather than unjustified*, not what with what is required for one to have the belief in the first place. The latter is a requirement for the existence of a *candidate* for justification; the requirements for justification become relevant only after those requirements have been met.

The above characterization of immediate justification is purely negative; it is justification by something that does not include the possession of reasons or evidence. This leaves room for a considerable diversity. In particular it stretches over both the following:

- (a) When a person acquires a perceptual belief (e.g., that there is a tree in front of him), his belief is justified provided it was acquired via the operation of his sense organs in "normal circumstances"
- (b) If a person clearly understands what " $2 + 2 = 4$ " means, he is thereby fully justified in accepting that proposition.⁶

The justification of perceptual beliefs as envisaged in (a) is *immediate* because what does the justifying does not include the believer's possession of suitably related *knowledge* or justified belief. The belief is justified provided it has originated in a certain way; it is not also required that the believer know, or even believe, that this is the case. Nevertheless in (a) there are conditions required for justification that go beyond what is otherwise required for knowledge. And in this respect it differs from (b). There it is claimed that I can't lack justification for believing that $2 + 2 = 4$, provided I understand what that means, and of course unless I understand what it means I can't even have the belief. So here nothing is required for justification over and above what is otherwise required for knowledge (in this case the belief condition). Immediate justification like (a), where there are "additional" conditions for justification we may term "heteronomous"; while immediate justification like (b), where nothing is required for justification over and above the other requirements for knowledge, we may term "autonomous."

Clearly self-warrant is a form of autonomous justification, since only the belief condition is required for justification.⁷ There is another form of autonomous justification that has received considerably more attention, that in which it is the satisfaction of the truth condition that does the justifying. Thus in *Theory of Knowledge*⁸ R. M. Chisholm writes that whenever "what justifies me in counting it as evident that *a* is *F*

⁶ In the interest of avoiding red herrings, let me make it explicit that I am not (in this paper) asserting either (a) or (b). I cite them as *examples of claims* that count as claims to immediate justification, just in order to illustrate the spread of that category.

⁷ The self-evidence of (some) *a priori* propositions, used above as the paradigm of autonomous justification, differs from the self-warrant of *B*'s in two ways. First, with self-evidence the requirement for justification is less—only the *understanding* of the proposition is required. This is clearly not sufficient for *B*'s. I can perfectly well understand the proposition that I am now sensing a purple hexagon, but I have no tendency to believe this proposition and would not be justified in doing so. Second, with self-evidence, whenever a belief is formed with *that* propositional object it is *ipso facto* justified, whatever the identity of the believer. But with *B*'s, self-warrant does not follow the proposition across believers. Hence we speak of self-evident *propositions* but self-warranted *beliefs*. However, it is still true that the above claim about " $2 + 2 = 4$ " implies that beliefs with that propositional object are self-warranted, on our definition of "self warrant."

⁸ Englewood Cliffs, N.J., 1966.

is simply the fact that *a* is *F*...we have encountered what is directly evident" (pp. 26-27). Let's call this form of justification "truth-warrant."⁹ It is weaker than self-warrant. If *all* beliefs of type *K* are warranted just by belonging to that type, the same will hold for all *true* beliefs of type *K*, but the latter does not entail the former.

Those who think one's knowledge of his sensations and thoughts derive from one's "awareness" of them are in effect taking *B*'s to receive *heteronomous* immediate justification.

We are not only aware of things but we are often aware of being aware of them. When I see the sun, I am often aware of my seeing the sun; thus "my seeing the sun" is an object with which I have acquaintance...Similarly we may be aware of our feeling pleasure or pain, and generally of the events which happen in our minds. This kind of acquaintance which may be called self-consciousness, is the source of all knowledge of mental things.¹⁰

Russell (and this is typical of partisans of "direct awareness") does not speak of beliefs in this passage. Nevertheless, for reasons analogous to those presented in Part IIA, Russell will have to admit that we do normally believe that we are having a certain feeling when we do have it. And if the question is raised as to what if anything justifies such a belief, Russell would surely have to say that it is justified by the fact that the person is aware of the feeling. Thus Russell is committed to the view that *B*'s are justified by one's direct awareness of the *C*'s in question. This will count as immediate justification provided it is not *required* that the subject know or justifiably believe that he is aware of the *C*, but only that the subject be aware of the *C*. And the justification will be *heteronomous*, since this requirement goes beyond the other requirements for knowledge. Even if it is held (and even if it is truly held) that one *cannot* have a *B* without also being directly aware of the *C* in question, it will still be the case that the requirement of the latter is distinguishable from the requirement of the former.

Part III

ARE *B*'S SELF-WARRANTED?

Now for my reasons for accepting (6). Since the negative part of that principle is quite uncontroversial I shall restrict myself to supporting the positive part (5).

In Part I, I pointed out that when a belief is justified it is by virtue of satisfying the requirements laid down by some general principle of justification. Thus in accepting (5) I am presupposing the following epistemic principle:

- (7) Every *B* is warranted just by virtue of being a *B*.

And here the question of whether *B*'s are self-warranted can be construed as a question about the status of this principle. As with other dimensions of evaluation, we may distinguish a *de facto* and a *de jure* version of our question. On the *de facto* interpretation we are asking whether (7) is in fact accepted in some "epistemic community" (usually our own). On the *de jure* interpretation we are asking whether (7) is validly or justifiably accepted.

The *de jure* question is, of course, the one in which we are ultimately interested. If (7) is the principle we should accept, it is of secondary importance how we do ordinarily proceed. I shall be taking the *de facto* investigation only as a background for the *de jure*.

As it is neither easy nor profitable to effect a rigid separation between the two questions, I shall proceed as follows. First, I shall present reasons for supposing that (7) is the principle we actually employ for *B*'s. Then I shall discuss an objection to that argument, viz., that we tacitly presuppose further requirements for justification. I shall consider two putative further requirements: *truth* and *immediate awareness*. But here rather than discuss the *de facto* issue head on, I shall sidestep it and argue that (a) immediate awareness cannot justify *B*'s (whatever we *might* ordinarily suppose), and (b) a truth-warrant principle is not acceptable if (7) is acceptable. Then I shall proceed to argue that (7) is the principle we *should* accept for *B*'s.

⁹In "Varieties of Privileged Access" I termed it "truth-sufficiency."

¹⁰Bertrand Russell, *Problems of Philosophy* (London, 1912), p. 77. See also G. E. Moore, "The Refutation of Idealism" in *Philosophical Studies* (London, 1922), pp. 24-25; "The Nature and Reality of Objects of Perception" in *ibid.*, pp. 70-71; *Some Main Problems of Philosophy* (London, 1953), p. 49. C. D. Broad, *The Mind and Its Place in Nature* (London, 1929), pp. 310-311.

But we will not have a thesis sufficiently determinate to be worth arguing about until we have given a more explicit indication of the boundaries of our *B*-category. To that task I now turn.

A. What Counts as a *B*?

'*B*' was initially explained (p. 3) as a belief of a certain person that he is currently in a certain *C*. This means that our concept of a *B* is no more determinate than our concept of a *C*; and so I must do something to indicate the boundaries of the latter.

First, the intuitive idea of a *conscious state* I mean to be employing is that of a "mode of consciousness" or "way of being conscious." *S*'s being in a certain *C* simply consists in *S*'s being conscious in a certain way; and *S* is in that *C* only for so long as he is conscious in that way. This is already sufficient to exclude "dispositional states" like wants, beliefs, and attitudes, which may be possessed during periods when the subject is completely unconscious, and also during periods when nothing relevant to the particular state in question is in his consciousness. Whereas a (conscious) thought, sensation, imaging, or feeling is possessed only for so long as the subject is conscious in the appropriate way. I use the term "state" in a relaxed fashion to cover occurrences, activities, and processes, as well as states more properly so-called.

I should also explain just what is meant by "conscious" in "way of being conscious." However, this being one of the more intractable stretches of the conceptual terrain, I cannot go into the matter properly and still leave room for the epistemological issues that are the central focus of the paper. I can only indicate that my preference would be to explain "being conscious" in terms of the notion of "immediate awareness" that will be adumbrated in Part III, C, 2; so that a *C*, a way of being conscious, is a state that involves an immediate awareness of some datum. No doubt much needs to be said as to the kinds of data that are involved in conscious sensations, feelings, thoughts, and so on; but we shall have to forego that for now. I shall assume that even this much explanation will suffice for a wide area of agreement on particular cases.

But there is still a difficulty. Some beliefs that clearly satisfy the above criterion for *B*-hood should not be counted as self-warranted. To take an extreme example, suppose I believe that

I am currently having a feeling just like the one you had 47 days ago at this time; this is a belief of mine to the effect that I am currently in a certain conscious state, but no one would be tempted to regard it as self-warranted. If warranted at all, part of that warrant comes from whatever reasons I have for thinking your state of feeling at that time was of a certain character.

The general point illustrated by this counterexample is that a *C*, like anything else, can be identified, or conceptualized, in many different ways; and it depends on which conceptualization we use whether or not we have a strong candidate for self-warrant. Let us survey some modes of conceptualization that prevent a self-attribution of a *C* from being self-warranted.

- i. The type exemplified by the above case can be specified as: similarities between the state in question and others not possessed by the believer at the moment. This includes his own past experiences as well as those of others.
- ii. Causal diagnoses, e.g., "I have a toothache," where part of what that means is: "I feel a pain that is due to the stimulation of a pain receptor in one of my teeth."
- iii. *De re* conceptualizations of cognitions, e.g., my belief that I am thinking of *the mayor* that he wears loud clothes, where the most apt verbalization of the content of my thought is "The man who lives directly across the street from me wears loud clothes," and where I also believe that the man who lives directly across the street is the mayor. Since to be justified in the original belief I would have to be justified in the belief last mentioned, the former is not self-warranted. To avoid these difficulties we shall have to restrict *B*'s concerning thoughts to *de dicto* conceptualizations, which characterize referents in the same way as the thought in question.
- iv. Descriptions of images in terms of conformity with some external object, e.g., a belief that I now have the Rouen Cathedral before the mind's eye. My justification for this would seem to depend on whatever justification I have for my beliefs as to what the Rouen Cathedral looks like. To make the belief a strong candidate for self-warrant the image will have to be

described in terms of sensible qualities or, alternatively, in terms of what I take it to be an image of.

The common character of these restrictions may be put by saying that *B*'s are restricted to conceptualizations of *C*'s in terms of their "intrinsic content" or "intrinsic character," rather than in terms of external relations in which they or their contents stand, relations of causality, similarity, coincidence, or conformity. In putting the matter in this way we are assuming that we can distinguish the intrinsic features of, e.g., a sense-datum or mental image from its similarity to some actual paradigm.

This restriction to intrinsic content will also rule out other candidates for *B*-hood, where the difference between that candidate and the nearest self-warranted belief is of a somewhat different character. In the above cases it was a matter of different conceptualizations of what it is natural to regard as one and the same state; whereas in those cases to which we now turn the nearest *B*-correlate concerns a *C*-component of the original state, rather than that state differently conceptualized. Thus cognitive reports with an external "success implication" like:

- (8) I see a beech out there.
- (9) I just remembered that Jack said he was coming at 5:00.
- (10) I know that you won't like it.

entail that certain states of affairs obtain over and above the present consciousness of the reporter. (8) entails that there is a beech out there, (9) that Jack did say that he was coming at 5:00, and (10) that you won't like it. Hence a belief that (8), (9), or (10) clearly is not self-warranted. These beliefs are best regarded as beliefs that one is in a certain *C*, plus a further claim as to its correspondence with external fact. To make them bona-fide *B*'s, we simply strip away that latter component, thus transforming our initial reports into:

- (11) It appears to me visually that there is a beech out there.
- (12) I just had what I take to be a recollection that Jack said he is coming at 5:00.
- (13) I feel certain that you will not like it.

A somewhat analogous, though more complicated, reduction will have to be performed on full-blooded emotional state concepts to render them usable in *B*'s. Thus my belief that I *am* indignant at Nixon, will have to be pared down to the belief that I *feel* indignant at Nixon before it can be admitted as a *B*.¹¹

These considerations show that '*B*' is to be defined in terms of a kind of concept rather than a kind of state, more specifically, a concept of the intrinsic character of some *C*. So our chain of definitions should go as follows:

- (14) *C*—a way of being conscious.
- (15) *I*-concept—a concept of the intrinsic nature of some *C*.
- (16) *B*—a belief of a person that he is currently in a certain *C*, where that *C* is conceptualized by an *I*-concept.

B. Do We Ordinarily Accept a Self-Warrant Principle?

The following remarks may serve as a background to a consideration of the *de facto* question. I take it that the members of an "epistemic community" judge beliefs to be warranted or unwarranted according as they note certain conditions to be or not to be satisfied, and that there is a considerable consistency among members of the community as to what conditions are (in practice) required. Of course in order to be a "practicing member" of an epistemic community, one need not have explicitly formulated the principles (rules, norms, standards), that lay down the requirements for a certain kind of belief, any more than a member of a linguistic community has to have formulated phonological rules in order to make judgments of correct or incorrect pronunciation in agreement with other members of the community. By observing the conditions that evoke judgments of "warranted" or "unwarranted" for beliefs of a certain kind, an observer can discover what epistemic rules are being implicitly employed in the community; and a reflective member of the community can disengage the rules by reflecting on the conditions under which he would make one or the other judgment.

My defense of the claim that (7) is accepted in our epistemic community consists of con-

¹¹ For some indication of what is involved see my "Emotion and Feeling" in Paul Edwards (ed.), *Encyclopedia of Philosophy* (New York, 1967), Vol. II, pp. 479-486.

siderations of the sort last mentioned. In reflecting on our epistemic practice I note that we require nothing for the justification of a *B* over and above its being a *B*. If *S* tells us that he has the image of a snow-capped mountain before his mind's eye, that he feels relieved, or that he is thinking about the election, then, provided we are convinced that his words do straightforwardly express his beliefs, we take it without more ado that he is fully warranted in holding those beliefs. We would unhesitatingly brand as absurd a request for justification such as "Why do you believe that?", "What reason do you have for supposing that?" or "How do you know that?"

C. Do We Assume Other Requirements?

To this argument it may be retorted that although we do customarily proceed in the way just described, that is not because we take *B*'s to be self-warranted; rather it is because we tacitly take the extra requirements for justification to be so obviously satisfied as not to require mention. Now that is a common phenomenon. We often assume without question that a person is justified in a belief (e.g., that he lives in Dubuque) that is *not* regarded as self-warranted. But in those cases we can readily specify what the requirements are that we unhesitatingly assume to be satisfied, e.g., that the person has often formed reliable perceptual beliefs concerning his place of dwelling and has not suffered an abnormal loss of memory. While in the present case the common man would be at a loss to answer "How do you know?", except by saying something like, "Well, it's my mental image, isn't it?", i.e., by simply repeating what is already implied by the original statement. Thus if our unquestioning endorsement of every *B* is due to the assumption that some further requirement is invariably satisfied, this assumption is beyond the reach of everyday reflection and requires some special digging to bring it to light. Let's consider the more prominent possibilities.

1. Truth as an extra requirement

A partisan of truth-warrant, like Chisholm, will hold that we employ the following principle:

(17) A *B* is warranted if and only if it is true.

Hence he will take it that we accept all *B*'s without question only because we take the extra requirement of truth to be invariably satisfied.¹² Is this the case? That is very difficult to determine. First it is difficult to decide whether we do believe that *B*'s are invariably correct. Of course if *B*'s are logically infallible or indubitable then we do believe them all to be true insofar as we are unconfused; but these are dubious claims. Let's suppose, for the sake of argument, that we do suppose *B*'s to be invariably correct; we still are faced with the question whether we *require* correctness for justification, i.e., whether our acceptance of all *B*'s as warranted is based on that belief of invariable correctness. The most direct way of settling this would be to consider a hypothetical case in which we either became convinced of the incorrectness of a *B*, or at least came to *doubt* its truth, and then consider whether we would still take it to be justified. But in attempting to construct such a case, we would be led into highly controversial issues. For present purposes I can bypass all this as follows. Even if we do in fact (tacitly) employ a truth-warrant principle, we would not be justified in doing so, if we were justified in accepting (7). For since (17) differs from (7) only in making an extra requirement, if we are justified in taking *B*'s to be warranted as such, we are not justified in imposing additional conditions. Thus, whatever is to be said about the *de facto* status of (17), it is not justifiably accepted provided we can justify (7), as I shall go on to argue.

2. Immediate Awareness as an Extra Requirement

The other main candidate for "extra condition" is "immediate awareness." Earlier we saw that many theorists (at least implicitly) advocate a form of heteronomous justification for *B*'s, according to which a *B* is justified *iff* the believer is "immediately aware" of the *C* in question. This suggestion does have a certain plausibility. It is natural to think of ourselves as directly aware of our sensations, images, and feelings. And in unquestioningly accepting others' reports of their *C*'s, are we not tacitly assuming that a similar situation obtains with them? Consider

¹² That is, he will take this position provided he does admit that we unquestioningly accept all *B*'s as justified, and provided he holds that truth is the criterion of justification we actually employ for *B*'s, as well as the criterion we should employ. I am not concerned here to determine whether the historical Chisholm holds these views.

also the distinction between *C*'s and states of persons with respect to which we do not so unquestioningly accept their reports—e.g., the conditions of their organs and muscles. It is tempting to think that the reason for this difference is that whereas *C*'s are "presented" directly to one's awareness, these physiological conditions are not.

Before yielding to the temptation let's try to be more explicit about the concept of *immediate awareness*. It is often taken to stand in no need of explanation. But the persistence of debates as to what is and is not an object of direct awareness, and the persistent obscurity as to just how an immediate awareness of *x* supports beliefs about *x*, reveals the danger of taking the concept for granted. The first step toward clarification is to distinguish between awareness of particulars (this color-patch or that noise or that visual image) and awareness of facts (that this patch is red or that I am aware of an image of a sunny beach). Clearly the way in which the awareness figures in justification will differ for these types. If my awareness of a certain region of the visual field (first type) is adduced as the justification for my belief that the region is blue, we have to consider the question just how my awareness confers a warrant on some particular belief about that object? Does it confer a like warrant on every belief about that object, and if so how? And if not, what gives certain ascriptions their special status? In other words, we have a "gap" between object and conceptualization (or "propositionalization") to be overcome. But this problem cannot arise if it is the awareness of *the fact that this patch is blue* that is said to justify my belief that this patch is blue. Here the object of the awareness fits the belief like a glove. The belief is simply, so to say, the "recording" or "storing" of the object of awareness. But this advantage is purchased at a price, for there are serious questions as to the intelligibility of this latter mode of awareness, at least as figuring in justification. Can a fact be *presented* or *given* to a passive apprehension? In order to have a fact as object, is it not necessary to "work it up" through conceptualization, in such a way as to make its apprehension indistinguishable from a judgment or belief? In the latter case, the "immediate awareness" of the fact will turn out to

be just a prior true justified belief; so that the problem of what does the justifying, and how, will have broken out afresh. Because of these difficulties I shall restrict myself here to awareness of particulars. I shall take this to be logically independent of any propositional knowledge or belief about the object, and even of any conceptualization of it or capacity to conceptualize it.

There is a circle of terms about which philosophers are wont to move when explaining any one of them. Thus in "Knowledge by Acquaintance and Knowledge by Description,"¹³ Russell explains his technical term "acquaintance" as follows:

I say that I am *acquainted* with an object when I have a direct cognitive relation to that object, i.e., when I am *directly aware* of the object itself. When I speak of a cognitive relation here, I do not mean the sort of relation which constitutes judgement, but the sort which constitutes presentation (pp. 209–210).

Other terms brought into the circle by various writers include "directly (immediately) conscious of," "directly (immediately) present to mind (consciousness)," "directly (immediately) apprehend," "given" "displayed" "directly (immediately) experience," and "intuit." Unfortunately none of these terms give us more guidance than the others in understanding how direct awareness of *x* justifies beliefs about *x*. Hence by defining one in terms of some of the others we are not making any real progress.

The concept is delimited somewhat by those who stipulate that beliefs justified by immediate awareness are infallible, indubitable, and/or incorrigible. This restriction is often implied by arguing with respect to, e.g., physical objects, that they are not "given," since we can be mistaken about them.¹⁴ But even if, as I should deny, *B*'s enjoy such strong epistemic immunities, this requirement would be at most a peripheral element in the concept. Since it makes sense to attribute infallibility to beliefs that cannot be regarded as resting on immediate awareness of particulars, e.g., simple arithmetical beliefs, it can't be that immediate awareness of *x* consists of rendering beliefs about *x* infallible.

¹³ In *Mysticism and Logic* (London, 1918).

¹⁴ See, e.g., C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle, Ill., 1946), pp. 182–183; H. H. Price, *Perception* (London, 1932), p. 3.

At this point the wisest of our predecessors in this inquiry have despaired of finding a helpful verbal definition and have resorted instead to an "ostensive" explanation in terms of examples. Thus in *Some Main Problems of Philosophy*, G. E. Moore writes:

I want you to realize as clearly as possible what sort of thing *this* way of perceiving which I call 'direct apprehension' is. It is . . . that which happens when you actually see any colour, when you actually hear any sound, when you actually feel the so called 'sensation' of heat, as when you put your hand close to a fire; when you actually smell a smell; when you feel the so-called 'sensation' of hardness, in pressing against a table, or when you feel the pain of toothache, etc., etc. In all these cases you directly apprehend the sense-datum in question (p. 46).¹⁵

Quite a variety of examples are given by one or another writer, but the clearest ones, those in which it seems most unexceptionable to think of items being "presented" or "given," are those in which the objects are sense-data, like colored shapes and sounds, or mental images. Let's take the concept as fixed by this sort of example.

Now what about the idea that *B*'s are to be justified by immediate awareness of the relevant *C*'s? (On this topic I shall, as I indicated earlier, address myself directly to the *de jure* issue, arguing that "immediate awareness" cannot be what justifies *B*'s, whatever we might ordinarily suppose.) The trouble with that suggestion is that in the paradigm by which the concept of direct awareness is explained the object of the awareness was not a psychological state, but rather the *object* of a psychological state. Moore explained "direct apprehension" as that relation which one has to a color or sound, when "actually" seeing or hearing it. In this paradigm the "psychological state" is the awareness rather than the object of the awareness. The psychological state a *B* is a "belief in" is the *sensing* of the color or sound, the "entertaining" or "having" of the mental image, rather than the color, sound, or image itself. And this paradigm provides the only explanation of the concept we are able to find. Now of course this by no means *implies* that there cannot be immediate aware-

ness of sensings and imagings, as well as of their objects. Even if a sensing is an immediate awareness, there might also be an awareness of that awareness. Indeed both Russell and Moore insist that this is the case.¹⁶ However in this they seem to me to be mistaken. I cannot see that the operation of sensing, imagining, thinking, or feeling is "given" or "presented" to us, "spread out" before our awareness, in the way colors, sounds, and mental images are. But even if we should convince ourselves that we are immediately aware of our own sensings and imagings, an analogous question would arise at the next level. One's awareness that he is currently sensing so-and-so is itself a *C*, and so we have the problem of what justifies one in the *B* that consists in the belief that one has that *C*. If that again is an immediate awareness of the *C*, we are faced with the spectre of an infinite hierarchy of awarenesses. To avoid that we shall have to recognize at some point a *B* that is not justified by an immediate awareness of the *C* in question. Thus even if, as I should deny, first level *C*'s are presented to our awareness, it cannot be generally true that what justifies a *B* is immediate awareness of the *C* involved.

D. Are We Justified in Accepting the Self-Warrant Principle?

The time has come for the *de jure* question. We have already raised that question with respect to the alternatives to self-warrant. We have seen that *B*'s cannot be justified by immediate awareness of the *C*'s involved, and that if we are justified in accepting a self-warrant principle, the more restrictive truth-warrant principle is not acceptable. Hence at this stage the *de jure* question boils down to: "Are we justified in accepting (7)?" That depends, of course, on what it takes to justify a principle of justification. This methodological issue is little explored and bristles with difficulties, but I shall at least have to stick my toes in the water.

Let's begin by considering the role of the justification condition in attribution of knowledge. Why do we need such a condition? Why isn't true belief sufficient? Isn't Truth what we are really after in the believing game? The

¹⁵ See also G. E. Moore, "The Subject Matter of Psychology," *Proceedings of the Aristotelian Society*, Vol. 10 (1909-1910), pp. 36-62; "The Nature and Reality of Objects of Perception," *op. cit.*, p. 68. C. D. Broad, *op. cit.*, p. 145. A. J. Ayer, *The Foundations of Empirical Knowledge* (London, 1940), p. 61.

¹⁶ See the references to Russell and Moore in fn. 10.

answer, of course, is that although truth is quite enough along the dimension of the relations of our beliefs to what they are "about," the concept of knowledge is a multi-dimensional one. The justification condition is there because in recognizing that *S* knows that *p*, we are not only judging his belief that *p* to accurately reflect the way things are (the things he is talking about) but also saying something about his *position* vis-a-vis that range of fact. But what sort of thing about his position? To answer that question we need to complement the above insistence on the distinction between the truth and justification condition, with an equal insistence on their interrelation. Truth is still the object of the believing game, and the "positions" to be applauded in that game are those that are calculated to lead to the truth. That is, to be justified a belief must be reliable, must be at least highly likely to be correct. To spell this out just a bit, there must be relevant facts about the way in which, or the circumstances in which, the belief is acquired and/or held, such that given those facts it is at least highly likely that a belief of that sort will be correct. We may abbreviate this as follows.

(18) A belief is justified *iff* it is reliable.

This formulation cries out for further specification in more than one respect. Do the facts in question have to concern the way the belief was *acquired*, what keeps it from being dropped, both, or either? If other sorts of "circumstances" are relevant, what sorts? Is it enough that given those facts, the belief is *highly likely* to be correct or must there be a logical or nomological *impossibility* of its being mistaken? And if the former, what concept of probability is relevant here? I shall take it that there is some way of answering these questions that renders (18) both specific and adequate. As we shall see, even in this unspecific form the principle will serve in the support of (7).

Considerations of reliability and mode of belief formation have recently been stressed by epistemologists like Alvin Goldman¹⁷ and David Armstrong,¹⁸ who have opposed the idea that knowledge is to be analysed even in part in terms of justification. They have contended that

what has to be added to true belief to make it knowledge is reliability, rather than justification. Now I agree with these thinkers that reliability is required to distinguish knowledge from (mere) true belief; but on my version reliability has this efficacy just because it is what is needed for justification. Thus the opposition between, e.g., Armstrong and myself (apart from the details of the notion of reliability—a major exception!) has to do with the question of whether reliability functions as a straightforward criterion of knowledge, or whether it so functions only because it is a criterion for a certain kind of evaluation, a positive form of which is required for knowledge. This is like the opposition between naturalists who want to identify the criteria and meaning of ethical terms and anti-naturalists who want to keep them distinct.

What bearing do these considerations have on the question of how principles of justification are to be evaluated? We might take them to show that (18) is the sole principle of justification. If we took this tack we would be practicing sharp economy in our general principles, packing all the diversity into applications. However that is not our only choice. We can, and I shall, opt for an analogue of rule utilitarianism, rather than for an analogue of act utilitarianism. That is, we can take reliability as a criterion for more specific principles governing specific types of beliefs. On such a criterion a principle that lays down a certain set of conditions, *I*, as sufficient for the justification of beliefs of type *A*, is acceptable only if the following requirement is met:

(19) *A*'s that satisfy *I* are reliable.

Are there other constraints a principle of justification must satisfy? I suggest that there are two. First,

(20) *I* must be applicable.

What that means is that we can actually use these conditions for determining whether a given *A* is justified.¹⁹ This does not mean that we can infallibly determine this, nor that we can determine it in every case. But it does mean that by and large we can determine satisfaction with a high degree of reliability and without too much

¹⁷ "A Causal Theory of Knowing," *The Journal of Philosophy*, vol. 64 (1967), pp. 357-372.

¹⁸ *Belief, Truth and Knowledge* (Cambridge, 1973).

¹⁹ Cf. R. M. Chisholm's concept of a "mark of evidence" in *Perceiving* (Ithaca, N.Y., 1957), pp. 34-37).

difficulty. After all, evaluation of beliefs is a practical affair (albeit of a theoretical sort), and putative criteria that cannot be applied are thereby disqualified.

A third constraint is suggested by the following consideration. Let's suppose that perceptual beliefs concerning the physical environment are correct in the overwhelming majority of cases; and let's further suppose that this confers a high probability of correctness on every perceptual belief. Should we then adopt a principle that certifies every perceptual belief as such without further requirement? This principle would satisfy the first two requirements. Nevertheless it is not acceptable, and just because we are able to formulate further applicable conditions that increase materially the proportion of true beliefs in the approved class. That is, by adding conditions concerning the state of the perceiver and the environment and concerning the perceiver's orientation vis-a-vis the object, we are able to wash out almost all the false beliefs. Clearly where this is possible we must prefer the enriched set of conditions. We want to increase the proportion of true beliefs as much as possible, without sacrificing applicability. Hence our third requirement:

- (21) There is no other applicable set of conditions. I_1 such that the proportion of true beliefs is significantly greater among A 's that satisfy I_1 than among A 's that satisfy I .

Does (7) satisfy these constraints? As for (19), we have of course very little idea as to how B 's are formed. We have seen that the traditional "immediate awareness" story collapses on inspection, and we have no real replacement, apart from some speculative cybernetic models. However that does not prevent our being justified in supposing (7) to satisfy (19). If we have adequate reason for taking B 's (without further condition) to be always, or almost always, correct, then we are justified in supposing that B 's are always, or almost always, formed in such a way as to be reliable, even if we have no idea of what that way is.

Are B 's always, or almost always, correct? Of course if they are infallible, logically or nomo-

logically, they are all guaranteed to be correct. Even if they are indubitable or incorrigible, no contrary evidence can ever prevail against the subject's sincere belief, and so we will be debarred from having any reason for ever thinking any of them incorrect. But since we are assuming these strong immunities to be inapplicable, how can we justify the claim?

At this point I would simply allude to the extreme difficulty of finding *actual* cases in which a B is shown wrong. This difficulty is partly due to the absence of independent checks on the accuracy of B 's; to a large extent we have no alternative to taking the believer's word for it. But that is not the whole story. We do have a variety of ways of "getting at" the mental states of others, over and above their direct testimony. The extent and force of these ways vary from one kind of mental state to another. They are greatest with "dispositional" states that do not fall within our C -category. We can, e.g., get some purchase on what a person wants and believes by observing his goal directed behavior. If he walks to his car, gets in the driver's seat, inserts the key in the ignition, starts the engine, and drives away, this surely gives us a considerable basis for supposing that he wanted to transport his car to some other location; that he believed that inserting the key in the ignition was necessary for starting the engine, and so on. With C 's the independent indications are weaker but still not insignificant. We expect a person who feels depressed, relieved, or excited to betray this in a variety of ways in his look, bearing and behavior. And by listening to a person's conversation we can draw reliable inferences as to the thoughts that are going through his mind, even if he does not explicitly tell us that he is thinking ——. I would suggest that when we do employ these independent checks, we rarely find them at variance with the subject's report; where they are, the rarity of this will justify us in either opting for the self-report, or putting its falsity down to insincerity rather than mistake. Thus insofar as we have independent empirical evidence, we have no reason to think that B 's are ever mistaken.²⁰ I take this to be strong support for the claim that for all B 's some highly reliable mode of belief formation is operative, even

²⁰ This is a stronger conclusion that we need, and I do not wish to insist on it in this paper. For present purposes it will suffice to show that B 's are, so far as we can tell, rarely mistaken. And the extreme difficulty of finding convincing examples of mistake should be evidence enough for *that*.

though we are unable to say just what that is. And this means that (7) passes the first test.

As for the other requirements, (20) poses no problem. We are certainly able to determine whether the conditions are satisfied, since they simply consist in the beliefs in question being a *B*. If we have a belief before us for evaluation we can certainly tell whether it is a *B* or not. As for (21), if, so far as we can tell, *B*'s are never mistaken, we are quite without motivation for seeking extra conditions. And even if *B*'s are occasionally mistaken, there would seem to be no applicable conditions that would reduce the proportion of errors. The only remotely plausible candidate I can think of is that the subject, *S*, be in possession of his faculties, not be too confused, distracted, or whatever, to "know what is going on". But if this condition were violated to such an extent as to engender a significant possibility of *S*'s not being able to tell what he is thinking or seeing, he would not be in a position as much as to form a belief as to what he is thinking or seeing. But in that case there is no belief to be justified or unjustified, and so adding this condition does not rule out any mistaken *B*'s. I conclude that we are amply justified in accepting (7).

Part IV

TWO OBJECTIONS TO THE SELF-WARRANT PRINCIPLE

In the remainder of the paper I shall consider two objections to the self-warrant thesis.²¹ I discuss these objections not only because they raise difficulties that deserve to be met, but also because in the course of meeting them we are forced to make the thesis more explicit and to uncover various interesting epistemological points.

A. Are Unwarranted *B*'s Impossible?

The first objection is directed against the modality involved in the concept of self-warrant. (Remember that to say that *B*'s are self-warranted is to imply that it is impossible for a belief to be

a *B* and not be warranted.) The objection simply is that we can give consistent descriptions of states of affairs (there are possible worlds) in which (some) *B*'s are not warranted; hence even if, as a matter of fact, all such beliefs *are* warranted it cannot be *impossible* that such beliefs not be warranted; they are not necessarily warranted.

The descriptions most commonly proffered are of the following sort. Let us suppose that physiological psychology has developed to the point at which laws have been firmly established, linking types of conscious states to their neurophysiological correlates. Let us further suppose that the associated technology has been so developed that electronic devices can be used to determine, in fine detail, what is going on in the brain. In that case we might very well use such information to override the subject's sincere testimony as to what he is feeling or thinking. Note that so far we have an argument against incorrigibility but not against self-warrant. Even if it is possible for a person to be shown wrong in a *B*, it does not follow that *B*'s are not self-warranting. However our description can be extended so as to involve cases of unwarranted *B*'s. Suppose we identify enough cases of mistaken *B*'s to enable us to establish generalizations as to the conditions under which *B*'s are likely to be mistaken. For example, it might be discovered that people are always correct in their *B*'s when the overall pattern of brain waves is in the range W_1 - W_{49} , while if the pattern is a W_{50} , mistaken *B*'s occur frequently. After this discovery we would justifiably take a *B* to be warranted only if *S* were in some state of the range W_1 - W_{49} . Now something is required for a *B*'s being warranted other than just being the belief it is; and when a person has a *B* without meeting those extra conditions, his belief is unwarranted.

We can maintain (5), while admitting the possibility of this kind of development,²² by making more explicit the order of modality involved in the concept of self-warrant (where "orders" of modality comprise, e.g., logical and casual). The objection assumes that our formulation was in terms of logical impossibility: if that

²¹ Arguments of the sort I will be considering have rarely, if ever, been directed against self-warrant as such, since that mode of privileged access has been largely ignored in the literature. Rather they have been directed against one or another of the stronger forms.

²² Many philosophers would deny that it is possible. Thorny issues are involved, and I do not have time to argue the matter. I shall have to content myself with showing that even if the possibility be admitted the self-warrant thesis can still be saved.

were the case, then the exhibition of a consistently describable state of affairs in which one's belief that he feels relieved is not warranted suffices to refute the claim that it is *impossible* for any *B* to be unwarranted. However it is not necessary to employ *logical* modalities in order to have a distinctive concept of self-warrant. What seems to me the most natural way to understand impossibility of non-warrant, even apart from the necessity for meeting the present objection, is in terms of what we may call "normative" modalities. This is the kind of modality that consists in something's being necessary, possible, or impossible, relative to some valid norm, rule or standard. It is exemplified by such statements as:

- (22) It *can't* be right to deliberately take another person's life.
- (23) You *must* stand behind the baseline when serving.
- (24) You *can't* use a singular form of the verb with a plural subject.

To say something is *normatively impossible*, as in (22) and (24) above, is, of course, to say that it violates some norm, rule or standard, but that is not the whole of its meaning. Something can be in violation of a standard without being properly termed (normatively) impossible. For example, watching television on a certain occasion might violate some standard, as it would if I had promised to visit someone at the hospital during that time; but we would not say (at least on those grounds) that it cannot be right to watch television. We use the modal formulation when the lack of conformity is due just to the item in question being as described, without any further conditions required. In other words, we use the modal statement when the norm *by itself* implies the lack of conformity of the item in question, without the need for additional premises. In saying, "It can't be right to deliberately take the life of another human being," we are claiming that a valid (moral) rule forbids such actions explicitly and categorically, whatever the circumstances. Whereas my evening of television watching would be considered wrong not because any moral standard categorically forbids television watching as such, but because the moral rule that one should keep his promise, *plus* the facts that I had promised someone to visit him in the hospital that evening and watching television prevented my doing so, all to-

gether implies that it was wrong of me to watch television. This distinction between "isn't right" and "can't be right" is parallel to the distinction between "isn't so" and "can't be so" in casual and logical modalities. Thus a determinist, who holds that every event is causally determined to happen just as it does happen, can still make a distinction between what is causally impossible, and what simply *does* not happen though it is causally possible. The causally impossible, e.g., my jumping out of a window in still air and moving upward, is that the non-occurrence of which can be deduced from causal laws alone, without the need of additional premises. The causally possible but in fact non-existent, e.g., my skiing at this moment in Switzerland, is that the non-occurrence of which can be derived from causal laws only with the aid of additional premises specifying particular matters of fact. Again, the *logically* impossible is that the falsity of which follows from logical truths alone; while the merely *de facto* false is that the falsity of which follows from logical truths only with the aid of some factually true premises.

Now, as we have pointed out, "warranted" is a normative or evaluative notion. To say that a belief is (epistemically) warranted is to say that the believer has an (epistemic) *right* to his belief, that in accepting it he is proceeding as he *ought* (epistemically), that it is (epistemically) *justifiable* for him to accept it, that the belief (or its acceptance) is in accordance with a valid epistemic *norm*. Hence when we speak of the *impossibility* or *necessity* of a belief's being warranted, it is natural to understand these modalities as being of the normative sort. To say that it is *impossible* for a *B* to be unwarranted is simply to claim that there is in force a valid epistemic norm (7), that categorically declares *every B* to be warranted, with no further conditions required. The warrantedness of any particular *B* follows from that rule alone without the necessity for additional premises.

On this interpretation can we hold that *B*'s are self-warranted while recognizing the *logical* possibility of a *B*'s not being warranted? Yes, because it is *logically* possible that we should (justifiably) employ a norm for *B*'s other than (7), a norm which would require further conditions for warrant. That possibility is, in fact, at the heart of the sort of counter-example adduced by the present objection. *If* such counter-examples are indeed logically possible, then they

exhibit a logical possibility of *B*'s not being warranted *just because* they involve radical changes in those features of our present condition (primarily our inability to be sure that *B*'s are ever mistaken) that render (7) justifiable. Hence in those circumstances other norms for *B*'s, requiring additional conditions for warrant, would be justified; and *this* is how unwarranted *B*'s would occur. But the logical possibility of all this is quite compatible with (7)'s being a valid epistemic norm in our present situation, and hence with its presently being *normatively impossible* for a *B* to be unwarranted. Thus on the normative modality interpretation II survives the objection.

B. Is the Validity of (7) Part of What Justifies a B?

We have been supposing that *B*'s are warranted only because (7) is a valid principle. But if so is it not the case that what warrants a *B* is not just its being a *B*, but that plus the fact that (7) is valid? In which case the *B* could hardly be termed *self-warranted*. Of course if the validity of (7) followed from the *concept* of a *B*, then the "extra" requirement would not really be extra, but since we have supposed this not to be the case, we cannot make use of that defense.

The trouble with this objection is that it involves a fatal confusion of roles. Where *x* justifies *y* (renders it acceptable, desirable, fitting, right or good) there is always a norm, standard, rule, or principle in the background, which, so to say, empowers *x* to do the justifying; it is essential to justification that there be adequate candidates for *both* these roles—that which justifies, and that norm or principle by virtue of which it is able to do so. These are different roles and they must be kept distinct. Without a valid principle in the background nothing can do the justifying, and of course the principle by itself is not enough without something to satisfy it. Thus, to return to the case at hand, where a

belief is justified just by virtue of being a *B*, being a *B* does suffice for justification only by virtue of the fact that (7) is a valid epistemic norm. But if we were to suppose that in this situation what justifies the belief is not just its being a *B*, but that plus the fact that (7) is valid, we would have confused these roles. To admit that (7) is valid is to admit that being a *B* is all that it takes to justify a *B*; for that is what (7) lays down as the sole requirement. Admitting the validity of (7) while also insisting that more is required for justification is self-contradictory.

This confusion is interestingly analogous to the confusion of premises with rules of inference exposed by Lewis Carroll in his delightful essay, "What the Tortoise Said to Achilles,"²⁸ and the same unhappy fate awaits it. If we insist on converting any rule of inference we employ into a premise, we are launched, as Carroll showed, onto an infinite regress; for each addition of a new premise defines a new argument, which employs a new rule of inference, which in turn becomes converted into a new premiss, which in turn . . . Similarly if we suppose that the validity of the presupposed norm is part of what justifies the belief, then the justification by *x* plus the norm presupposes a different, higher level norm, which in turn will be made part of the justification, which then presupposes still a higher level norm, which in turn . . . Thus on pain of an infinite regress we are forced to distinguish what justifies the belief from the norm by virtue of which it does so, and so are "forbidden" from including the latter in the former.

Part V

I take the considerations of this paper to have shown that the self-warranted status of *B*'s must find a place in any adequate epistemology. Just what that place will be depends on the other features of that epistemology.

Douglass College
Rutgers University

Received July 31, 1975

²⁸ *Mind*, vol. 4 (1895), pp. 278-280.

III. THE REALM OF THE MORAL

JAMES R. FLYNN

THE problem of justification has taken centre stage since World War II, particularly the problem of how humane men can justify their ideals against non-humane and elitist opponents. I wish to analyze three thinkers, S. E. Toulmin, Kai Nielsen, and Kurt Baier, all of whom share, it seems to me, a certain approach: they use delineation of the realm of the moral against their ideological opponents—they attempt, whether explicitly or implicitly, to exclude Nietzscheans, Social Darwinists, elitists in general from the realm of the moral.

To clarify this approach, I must emphasize the distinction between a criterion of the realm of the moral (which separates the moral from the non-moral) and a criterion of the morally right (which separates the morally right from the morally wrong). Most of us do not think of our ideological opponents as having no morality, rather we think they have a competing morality, a competing criterion of the morally right, albeit one we hold to be vicious and (if we think we can make a rational case against them) indefensible. In other words, we do not deny that a Social Darwinist is a resident of the realm of the moral along with ourselves, we merely condemn his moral principles in terms of our own. On the other hand, we might call a man non-moral if we discovered that he seemed to have no other-regarding principles at all, if he were a mere egoist or sadist. Further, we would not have to confine ourselves to labelling his principles, e.g., someone who looked at the world purely through spectacles of self-interest might be said to lack a moral *point of view*, or someone who never gave anything but his own interests as a reason for his actions might be said to have a non-moral *mode of reasoning*, and so forth.

This brings us to a complication. It is not clear that all three of our thinkers would rule their opponents' principles outside the realm of the moral. For example, Baier says that moral

principles can be true or false and I rather suspect he would classify Social Darwinist or Nietzschean principles as false moral principles rather than as non-moral.¹ However, it is quite clear that an elitist's point of view would be a non-moral one—Baier tells us that a man has a moral point of view only if he agrees that rules and laws must be for the good of everyone alike.² Moreover, this is all that counts, for Baier tests principles against what he calls "the moral point of view" and therefore, the elitist's principles are doomed to be "false." It is no favor to Nietzsche to let his principles into the realm of the moral simply to be slaughtered. One could say much the same of Nielsen: he may grant his opponents' principles entry, but he holds their "mode of reasoning" to be non-moral (and holds that they have misconstrued the "functions of moral discourse") and it is the latter which he uses against them in ideological debate.

In sum, whatever these thinkers hold to be most fundamental in ethics, whatever the problem of justifying oneself against one's opponents turns on, it is this that they rule out (of the realm of the moral) on behalf of the non-humane or elitist. They can afford to be generous about principles, someone's principles are not likely to last long once cut off from the point of view or mode of reasoning that lends them sustenance. Indeed, our three thinkers levy heavy penalties: Baier says his opponents have principles that are false or not genuine; Toulmin leaves them with no good reasons for their ideals or at best with non-moral reasons; Nielsen describes them as non-moral, only quasi-moral, or at best moral by way of some spurious factual claim. Certainly, if we can pin such labels on our opponents, we will have a tremendous advantage in competing for the allegiance of mankind. As Nielsen says at one point, given a debate over what course of action to follow, if we are committed to a moral point of view, the issue is "already decided" in

¹ Kurt Baier, "The Point of View of Morality," *Australasian Journal of Philosophy*, vol. 32 (1954), pp. 132-133.

² Kurt Baier, *The Moral Point of View* (Ithaca, 1958), pp. 207-208.

Toulmin's favour.³ That is, the audience must choose between abandoning our opponents and abandoning morality and (I would add) it is fair to assume that the overwhelming majority of mankind think of themselves as moral rather than non-moral.

Those who defend the realm of the moral approach often say that surely, it is legitimate to offer some account of what separates the moral from the non-moral. No doubt, but this begs the question of what sort of account we should offer. Few would object to distinguishing morality from art, science, logic, religion, custom, law, or etiquette—and few would object to labelling as non-moral those who have never claimed to have ideals which deserve to take precedence over individual pleasures or tastes, e.g., plain egoists (as distinct from invisible hand egoists) or sadists. Controversy begins when we fail to stop at a broad account of the realm of the moral and go on to a narrow one. By a *broad* account, I mean one that grants entry to all who have attained historical significance as ideologues, as competitors for the allegiance of mankind, whether humane or non-humane, whether elitist or equalitarian, whether they consider the average man precious or expendable. By a *narrow* account, I mean one which excludes those who do not hold a certain set of basic ideals, let us say humane-equalitarian ones. It comes down to whether we find a place for only Jefferson and Marx or also Nietzsche and the many descendants of William Graham Sumner.

This distinction made, and it must be remembered throughout, I can present my case. Like Luther, I have some theses to argue, although only four in number: (1) A narrow account of the realm of the moral must be shown to be non-partisan; (2) It is not clear that the above can be done; (3) Even if the above can be done, the realm of the moral approach will be proved useless in the process; (4) To be useful, it must be transformed.

* * *

Thesis One: A narrow account must be shown to be non-partisan.

Under this heading, I want to take a look at

Toulmin and Baier and call attention to a curious fact: although it is quite clear just who they are expelling from the realm of the moral, these outcasts are *never named*. For example, when Toulmin argues for his account of the realm of the moral, he never faces up to even one ideological opponent. His most detailed defence of his account occurs on pp. 160–162 of his book.⁴ He reminds us of his two rules or criteria of moral reasoning, assessing acts in terms of an accepted principle or social practice, assessing social practices in terms of “general fecundity,” presumably alleviating suffering with equity. As for the sort of opponent he has in mind, it is someone who might reject one of his criteria in favor of total reliance on the other! Those who ignore criterion two, men who never assess social practices (and elites) in favor of citing them, are said to rely purely on authority and are not reasoning morally at all. Those who ignore criterion one, men who never cite social practices in favor of a “universal test of consequences,” are accused of substituting expediency for morality. And then comes his conclusion: “Consequently, even if all we do is to give up one or other of our present logical criteria, we turn ethics into something other than it is. And if this is the case there is no need for us to go on and consider more drastic alternatives: they can be ruled out at once.” In other words, tampering with Toulmin's two criteria has proved so obviously mistaken, we can forget about our ideological opponents, those who might propose a radically different alternative, e.g., someone who might say that social practices are to be assessed in terms of rewarding superior men and keeping the masses at bay. Every now and then, there are signs of an unwelcome guest, but he is quickly dismissed; for example, we may sometimes be lax enough to “ignore the sufferings of ‘inferior’ people” (clearly a Social Darwinist might do this on principle) but we know “very well” that this is beyond the pale of moral reasoning.⁵

If one wishes to find any non-humane ideologues in Toulmin, one has to dig into his footnotes. Apparently Bertrand Russell told him that his account of the function of ethics would not have convinced Hitler. Toulmin interprets this as a contradictory demand that he produce

³ Kai Nielsen, “Good Reasons in Ethics,” *Theoria*, vol. 24 (1958), p. 27.

⁴ S. E. Toulmin, *The Place of Reason in Ethics* (Cambridge, 1950).

⁵ *Ibid.*, p. 164.

a reasonable argument which would convert an unreasonable man, a lunatic blind to logic and evidence.⁶ But certainly Russell meant to ask him if he had a case against a non-humane opponent, a rationally defensible case, setting aside whether or not such would actually convince a particular opponent. There is one other footnote which seems relevant: Toulmin imagines someone objecting that his criteria are not exhaustive—that two communities could be equally valuable in terms of promoting “the fulfilment of everyone’s aims and desires” but one could be preferred because its members enjoy poetry while their neighbors enjoy only things like parlor games. He replies that it would be sensible to prefer one on aesthetic grounds, but not on *moral* grounds.⁷ We must take considerable liberties with this to make it into a reply to a non-humane opponent, but it is important to note its implications: take an aesthetic elitist, someone who thought that a great cathedral was worth the sufferings of thousands of ordinary men; following Toulmin, we would have to tell him that he was using aesthetic criteria rather than moral criteria—and that for anyone reasoning morally the issue was already decided against him. However, this line of argument, implicit in Toulmin’s account, is quite unconvincing: our elitist would reply that while *knowing* what constitutes a work of art is an aesthetic matter, there is no reason why he cannot make the *achievement* of such things (at the cost of much suffering) a moral imperative.

Baier does no better than Toulmin. Throughout his book, he promises to give reasons for having a moral point of view at all,⁸ but in the relevant chapter, he argues only against the thesis that we can use self-interest as the ordering principle of a human society. He says this would lead to a Hobbesian state of nature and that we can easily see that such a world would be worse than one both moral and social.⁹ Now such an argument merely poses what I call “the mystery of the vanishing Nietzschean.” When Baier expels men of enlightened self-interest and Nietzscheans and Social Darwinists and elitists of all sorts from the realm of the moral, we

expect to find them waiting for us in the land of the non-moral. But all we find is the man of self-interest, the others have simply vanished into thin air, they seem to be neither within the realm of the moral nor out, as far as troubling to debate with them is concerned.

Baier claims that “...since it has just been shown that moral reasons are superior to reasons of self-interest, I have given a reason for being moral, for following moral reasons rather than any other.”¹⁰ This is plainly false. The assumption is that the only alternative to morality is self-interest: this may be true if we admit elitists as well as egalitarians to the realm of the moral; but by excluding the former Baier has *created* a whole range of alternatives to morality that go beyond self-interest. Certainly elitist ideologues are not men of enlightened self-interest: they have a “point of view” which puts one’s ideals ahead of one’s own personal interests, they sacrifice themselves on the altar of genius or the military virtues or art. And certainly, the act of banishing them has no magical effect, they cannot be men of principle the moment before and men of self-interest thereafter. And finally, it is far from clear that the non-humane ideologue suffers from the same weakness as the man of self-interest, there is considerable *prima facie* evidence that he can use his principles to organize a human society. Baier cannot have it both ways: he cannot deny that non-humane ideologues have a moral point of view; and then ignore them when he comes to arguing against alternatives to a moral point of view.

In articles since *The Moral Point of View*, I can find only one line of argument that may be relevant. Baier argues that we can discriminate between social orders in that all of them are not equally “acceptable.” If some members of a social order find it unacceptable, this will bring on ennui and alienation plus an increased need for sanctions and greater suffering when these are used. The best system is one based on the principle of equity, one which resolves conflicts so that the interests of all are served equally. Baier goes so far as to say that by morality we *mean* a social order “in so far as it purports to be acceptable from the point of view of all its

⁶ *Ibid.*, p. 165 and 165 n. 2.

⁷ *Ibid.*, p. 137 and 137 n. 1.

⁸ Baier, *The Moral Point of View*, *op. cit.*, p. 298.

⁹ *Ibid.*, pp. 308–315.

¹⁰ *Ibid.*, p. 310.

members."¹¹ As it stands, all of this does not come to much. An elitist is likely to consider some alienation a small price to pay to maintain his ideals and goals; and he minds using sanctions and suffering much less than we. One cannot refute him simply by fiat, by talking about what morality "means."

Baier might say that morality consists of giving men reasons, reasons for putting something ahead of their self-interest, reasons which show that their doing this is someone else's business.¹² In rebuttal, the elitist has plenty of reasons, e.g., "cathedrals should come ahead of the interests of ordinary men and it is my business to see that they do." Baier would probably object that morality requires showing all men that it is rational *for them* to put something ahead of their interests and certainly, some men do not desire cathedrals.¹³ Again in rebuttal, the solution Baier offers to this problem, the principle of equity, is subject to the very same objection; some men (elitists) do not desire a state of equity and would not think it rational to put it ahead of *their* interests (it may after all call for considerable sacrifices on their part). I should emphasize that it is I who have applied these arguments to the ideologues in question and their apparent weakness may stem from this. Baier himself says nothing of Social Darwinists or Nietzscheans.

However, an omission of this sort simply raises suspicions about the whole approach. It is not difficult to imagine what our ideological opponents, for example, a Social Darwinist, will say against us: "I am not going to tamely submit to being labelled non-moral. All of these attempts to 'delineate' the realm of the moral are not what they seem, they are purely partisan in nature, they are the creations of thinkers with a humane-equalitarian bias and designed to admit them into the realm of the moral while keeping

non-humane men out. I have little sympathy with this kind of game but I can easily learn to play. I can simply use my own ideals and do some 'delineating' myself and then, I think you will find that it is humane men who misconstrue the function of ethics (they don't see that its real function is to promote the survival of the strong rather than the weak), and who reason non-morally, and who lack a moral point of view. Concerning the 'curious' fact that these thinkers expel men like me without acknowledging our existence, there is nothing curious about it. They must get rid of us or the whole approach just will not work; but if they ever admit what they are doing, even to themselves, the partisan character of their criteria would become too obvious to be ignored."

When we turn to academics, most of whom are no doubt impeccably humane, we find that the above suspicion cuts across ideological lines, witness: Mackie—Toulmin sets out criteria which are merely "the dominant ones at present"; Sacksteder—accuses Toulmin of eliminating other modes of reasoning by "arbitrary fiat"; Dykstra—the criteria are really "criteria accepted among a certain class or groups of people, e.g., those at Cambridge"; Nakhnikian—Toulmin ties the evaluative meaning of ethical to a "particular criterion"; and Wadia—Toulmin is actually "expressing a moral preference."¹⁴ I must say that I can see no way of answering the Social Darwinist except to take him at his word, that is, ask him for his account of the realm of the moral, show it to be defective, and do the same with other non-humane ideologues. One cannot convert partisan criteria into non-partisan criteria merely by shoving them up to the level of metaethics—they merely become partisan metaethics.

This raises a problem of method and Nielsen alone attempts to provide a solution. The fol-

¹¹ Kurt Baier, "Ethical Pluralism and Moral Education" in C. M. Beck *et al.* (eds.), *Moral Education* (Toronto, 1971), pp. 109–110; also "Indoctrination and Justification" in J. F. Doyle (ed.), *Educational Judgments* (London, 1973), p. 87.

¹² Baier, "Indoctrination and Justification," *op. cit.*, pp. 84–87; also "Obligation: Political and Moral" in J. R. Pennock and J. W. Chapman (eds.), *Political and Legal Obligation* (New York, 1970), pp. 134–138.

¹³ See Baier's remarks on the flaws of utilitarianism: "Meaning and Morals" in Paul Kurtz (ed.), *Moral Problems in Contemporary Society* (Englewood Cliffs, 1969), pp. 45–46; and "Moral Obligation," *American Philosophical Quarterly*, vol. 3 (1966), pp. 221–222.

¹⁴ See: John Mackie, "The Place of Reason in Ethics," *Australasian Journal of Philosophy*, vol. 29 (1951), pp. 114–124 (see p. 116); William Sacksteder, book review of S. E. Toulmin, *An Examination of the Place of Reason in Ethics*, vol. 62 (1951–1952), pp. 217–219 (see p. 218); V. H. Dykstra, "The Place of Reason in Ethics," *Review of Metaphysics*, vol. 8 (1955), p. 465; George Nakhnikian, "An Explanation of Toulmin's Analytical Ethics," *The Philosophical Quarterly*, vol. 9 (1959), pp. 59–79 (see p. 67); and P. S. Wadia, "Professor Toulmin and 'The Function' of Ethics," *Philosophical Studies* (Ireland), vol. 14 (1965), p. 91.

lowing is something of my own creation, I have had to correlate three separate passages from his early articles and amplify them in the light of his later work. But I think it represents his views on how we can test (using his terminology) various accounts of the characteristic functions of moral discourse: (1) Take stock of certain usages and concepts which we all know pre-analytically to be central to ethics—usages such as recommending how men are to live, laying down rules for dealing with conflicting interests (or demands), allocating praise and blame—concepts such as universalizability, justice (or fairness), happiness, and sympathy; (2) Imagine a wide range of competing accounts, i.e., our own plus those of our ideological opponents; (3) See whether we can make a case that one of these (preferably our own) is uniquely viable—see whether we can show that all the other accounts conflict in some way with the above usages and concepts. This done we would have shown that an account of the realm of the moral can be *both* narrow and non-partisan, narrow enough to be used against our opponents and yet non-partisan in status.¹⁵

* * *

Thesis Two: It is not clear that we can show that competing accounts are non-viable.

In 1957, Nielsen published an important article, "The Functions of Moral Discourse," in which he sets out a number of accounts which might be considered alternatives to his own. Of these, there is only one I would offer if I were a non-humane ideologue:

The characteristic functions of moral discourse are to guide conduct and alter behavior so as to develop a superior class of man for which the rest of mankind are to exist simply as a means. That is, the rest of mankind are not to be regarded as moral agents with a worth of their own.¹⁶

As for what Nielsen has to say against this account, he makes three points: that to be a man is "by definition" to be a moral agent; that man is the sort of animal to whom moral praise and blame attach; and that any criterion we use to

give certain human beings moral priority over other human beings must be universalizable.¹⁷

Concerning Nielsen's first point, he says it has to do with the meaning we attach to the word "man"; and man as a moral agent is taken to mean that all mankind have "a worth of their own." If this point is really a verbal one, then the elitist has an easy way out: if he believes that a certain portion of homo sapiens are not worthy of regard, he need only stop calling them "men." Indeed, as Nielsen says, this is what elitists have done throughout history, the Greeks called their slaves barbarians, the Germans called the Jews submen, Nietzsche spoke of herd men, and so forth. On the other hand, if the point is not verbal, if it means "all members of the species homo sapiens *deserve* to be treated with respect," it is a substantive assertion—and subject to challenge by elitists who do not endorse its substance. Perhaps this is why Nielsen goes on to his second point, which has to do with what men deserve.

His second point is that man is a certain sort of animal, the sort capable of reflective choice it appears (except when he is an infant, senile, etc.), and that we must judge him in terms of praise and blame. This raises an important question, whether or not non-humane ideologues possess a standard of justice, for we would not call someone just if he allocated boons and ills to people without using praise and blame, or at least saying something about their merits. Let us give a Social Darwinist a chance to comment: "As for using praise and blame, I praise and blame men a good deal more vigorously than you do. I realize that you humanists are not necessarily crude equalitarians, that you recognize certain exceptions to equal treatment such as 'women and children first' and rewards for unusual merit. Nonetheless, you cling to the notion that just because someone is born into the species homo sapiens he has a right to a certain basic level of consideration. I rather think a man must pass a more demanding test than *that!*—he must earn even a minimal level of respect by proving himself tough, energetic, and intelligent. My ideal is the soldier who is also a tactician, a man who

¹⁵ See: "Good Reasons in Ethics," *op. cit.*, pp. 16-17; "Justification and Moral Reasoning," *Methodos*, vol. 8 (1957), pp. 99-100; and "The Functions of Moral Discourse," *The Philosophical Quarterly*, vol. 7 (1957), pp. 243-244.

¹⁶ *Ibid.*, p. 244.

¹⁷ *Ibid.*, p. 245.

proves his worth in a competition more subtle, and perhaps more demanding, than war. Those who fail are blameworthy all right, so blameworthy that they are worthless—they deserve punishment rather than compassion and some of them just deserve to go under. It is *because* I think of men as moral agents in terms of praise and blame that I deny they have any 'worth of their own'."

The humanist would hardly let the argument go at that and certainly Nielsen would not. In a recent work, he takes up the question of departures from equal treatment and says that such departures must be fair—one cannot allocate advantages and disadvantages to men "at least for the vast majority of cases, on grounds by which a person by his own good effort could not be in a position to achieve."¹⁸ The themes of fairness and justice as fairness are struck in a number of Nielsen's later works and they open up an obvious line of attack: that if the Social Darwinist has as his goal a never-ending process of "perfecting" mankind, he must weed out a certain number of the "unfit" from every generation—he must set a test that some men cannot pass no matter how hard they try; and that the Social Darwinist has never seemed particularly concerned with giving all men an equal chance to pass his test—he seems content to let the social environment favor certain children over others as they prepare themselves for the competition to come. In effect, Nielsen might accuse his opponent of cheating men of the right to *at least be judged* in terms of praise and blame, by condemning them for what they cannot help and therefore for what they cannot be blamed.

To give the Social Darwinist a right of reply: "There are a lot more people capable of transcending an unfavorable environment than you humanists like to admit—for every ten men who have not I can point to one who has, a man of average intelligence and talent who decided to stand and fight, and the existence of that one man implies that the other ten are blameworthy. Further, I would call the sort of environment you want for everyone just as crippling as an over-harsh environment, a pleasant and sentimental upbringing produces pacifists and humanists rather than men who are tough and competitive. However, I do not wish to be mis-

understood. Concerning the unfit, I know they suffer and I am not a sadist—I set no value on suffering for its own sake. If I had a magic wand that could tell me who the unfit are (in advance of their being tested), could give them bliss from the act of breathing, and could make sure they did not reproduce, I would use it. But pain exists, it is hard to eliminate, resources are limited, and further, thousands of things are more important than the elimination of suffering, things that fire the imagination, the race for space, the chance to rival the pyramids, giving our armies, and athletes, and industries all they need to compete successfully, and so forth.

"Concerning fairness, the same holds true. I know that some men, given their limitations, are blameless and I set no value on punishing *them*. But if one makes an idol of fairness, one ends up sacrificing too much that is important on its altar. The task of levelling all environments so as to give all men an equal chance is immense. And even then, we would not have begun to satisfy the appetite of fairness. It is unfair that some men suffer from what is not their fault, that one is lucky in marriage and another is not, that one is intelligent and another stupid—perhaps we should reserve the most desirable mates for the latter as partial compensation. It is unfair that some men are born deformed or mentally defective—perhaps we should render every child defective so that we will all begin life as equals. If we are really serious about fairness, we would have to deprive every man of every advantage he has not earned.

"My main point is this: justice as fairness is important, there can be no competition without rules and certainly we must be fair in establishing guilt in criminal cases; but justice as fairness is not everything. There is another sort of justice, justice as rewarding excellence, which is no less venerable and in my opinion, a good deal more important. Even our schools, which have been influenced far too much by humanists, give a mark for achievement as well as a mark for trying. If espousing justice as rewarding excellence puts one outside the realm of the moral, we had better start with Plato. I have every right to place fairness (particularly to the unfit) on a list of priorities and to place it well down the list. And when I say that some things are more important than others on this list, I mean to say

¹⁸ Kai Nielsen, *Ethics Without God* (London, 1973), pp. 82–83.

that it is a list of moral priorities. I feel a moral obligation to put things ahead of fairness, and a moral obligation to reward excellence, and a moral obligation to promote excellence. I have before my eyes a vision of the future, men strong, energetic, and intelligent locked in a competition in which the winners are admired and the losers are respected for asking no quarter."

A Nietzschean would want to add to the above: "The Social Darwinist has spoken about justice as fairness and justice as rewarding excellence and has ranked them in terms of importance. But he neglected something much more fundamental, he did not make explicit the criterion of importance he used to rank these two. I suspect his criterion is rather like mine, a *criterion of what constitutes a significant form of life*. Indeed, I challenge anyone to dispute this claim: however much he hides from it, however much he denies it, every thinker has such a criterion and builds his ethics on it. No doubt, there are a few who hold a truly equalitarian criterion, who consider the life of an ant just as important as that of a human being; but even most humanists are elitists, they think it right to sacrifice the interests of other animals to those of homo sapiens—the very phrase 'all men and their desires to be taken as of equal importance' implies that all *species* are not of equal importance. Now, I happen to have a criterion of importance, a criterion a living thing must meet in order to merit real consideration of its well-being, which most homo sapiens cannot pass. Look at the average man braying at a TV programme like 'This Is Your Life,' do you really think that fairness to such a creature even compares with furthering the existence, aims, and achievements of great men? In a word, fairness to superior men is important because *they* are important; fairness to others is not because they are not.

"Not that I am a Social Darwinist; he puts the 'soldier' ahead of the creative genius; and perhaps as a result, he is too much interested in perfecting mankind in general—why not be satisfied with a few great men and keep available a mass of mediocre men to serve as means to their ends? You may tell me that I underestimate the number of great men hidden amongst the masses and that if I improved their lot, this would give such men a chance to develop. This

argument appeals to me, for it implies that I have won you over—you are arguing in my terms, that any concern for the masses is justified solely out of fairness to the great. And finally, I hope no one is going to tell me I am mistaken about 'man'; I can describe man's behavior quite as well as you and can use the term in accord with any rules you care to lay down."

In the author's opinion, the Social Darwinist and the Nietzschean have made a good case (for inclusion in the realm of the moral), particularly impressive being the latter's contention about a criterion of what constitutes a significant form of life. Concerning the Social Darwinist's point, that justice as fairness is not everything, Nielsen himself eventually came to concede this (after a lapse of sixteen years):

But while justice is a central moral consideration, it is not the only relevant moral consideration and sometimes the claims of justice, where acting on them would cause great misery, should be set aside. In such circumstances, the innocent must suffer undeservedly because, unless they do suffer, a far greater total suffering will ensue.¹⁹

It is worth noting that Nielsen does not dodge the fact that he is putting an ideal ahead of justice as fairness, for humanists sometimes try to obscure this. They say that the above comes to no more than sacrificing fairness to a *few* men, so as to promote the happiness of *many* men or even mankind in general. In reply, this is not a matter of sacrificing fairness to a few on behalf of fairness to all—promoting happiness in general has (on the face of it) nothing to do with fairness. When humanists endorse such a goal, they make no direct reference to praise or blame, that is, they do nothing to show that mankind *deserves* happiness. In fact, it is hard to imagine any humanist worthy of the name who would calculate the comparative virtue of mankind at one time and another and then raise or lower the general level of suffering to match. The humanist ideal of alleviating suffering is something in addition to justice as fairness and each of us must ask himself just when the former takes priority over the latter.

There are some humanists, e.g., Elizabeth Anscombe, who would answer "never," but Nielsen is not one of these. And he objects strongly to the notion that he should be read out of the realm of the moral on these grounds,

¹⁹ *Ibid.*, p. 101.

that is, he thinks Anscombe (given her convictions) should say he has an "immoral morality," not that he has no morality at all.²⁰ I would add that if putting humane ideals ahead of fairness does not take one out of the realm of the moral, why should giving priority to non-humane ideals? And I take it that once the notion of putting non-humane ideals ahead of fairness is accepted, the way is clear not merely for the Social Darwinist (and the Nietzschean) but also for other non-humane ideologues, e.g., the aesthetic elitist, the man who would build one great cathedral at the price of undeserved suffering on the part of a thousand peasants.

We must take up Nielsen's third argument, that any criterion we use to give certain human beings priority over others must be *universalizable*. The rule of universalizability is complex, but in this context I take it to mean simply that if a man states a criterion of who deserves or merits preference, he must stand by it with logical consistency. Now some of our ideological opponents cannot pass this test. The racist is vulnerable because fundamentally, he faces the following dilemma: either he must assert correlations between color and personal traits (stupidity, laziness, etc.) in which case he can be routed by empirical evidence; or he must say that differences in color *alone* justify differential treatment of human beings. In the latter case, we can use logic to embarrass him in a number of ways. For example, and this is only one example, we can say: imagine you took a pill that turned you black but left you unchanged in every other way; do you really believe you should then go under, despite the fact that you remain the kind, intelligent, brave, witty fellow you are? Or imagine that everyone who is white turned black and vice versa; do you really believe that although they are totally unchanged in terms of life history, achievement, character, and personality, they should trade places?

However, a non-humane ideologue need not be a racist: the account of the functions of moral discourse Nielsen presents (on behalf of a Nietzschean) makes no reference to race; I certainly did not intend my Social Darwinist or Nietzschean to be taken for a racist. And this makes a crucial difference. We asked our racist to imagine himself going under with no change in his personal traits; but the above ideologues

state their criteria in terms of personal traits *themselves* and therefore, they need have no fear of questions about changing places with the unfit. When we ask a man who is strong and energetic and intelligent if he ought to go under if he were weak and lazy and stupid, it is really like asking him if he should go under if he were *somebody else*! We are asking him to imagine himself changed into someone radically different from the sort of man he is, changed into the kind of man he loathes. He is likely to answer: "Yes, I certainly should go under if I were like that, but do recall, I am not really like that at all."

Some thinkers believe they can win an easy victory over all elitists, non-racist as well as racist. However, they believe this because they link the rule of universalizability with one or more of the following: that one must have an equalitarian criterion of justice—accord all men a certain basic level of respect; that one must espouse justice as fairness rather than justice as rewarding excellence; that one must not only espouse justice as fairness but also give it priority over all other ideals. It is clearly a mistake to identify universalizability with these assertions or to think it logically entails such assertions. All of the above are *substantive* in character, they recommend one criterion of justice rather than another, or one kind of justice rather than another, or a certain criterion of importance; and universalizability is a purely *formal* rule, it must be supplied with a substantive principle before it can go to work (before it can demand that we stand by our principles with logical consistency). In my opinion, it is also a mistake to think that universalizability "suggests" equalitarian principles, the notion being that there is some sort of connection weaker than entailment. If an elitist can universalize his principles, it is hard to see why this will suggest anything non-elitist to him. He will claim that he stands by all three of his criteria, fairness, excellence, and importance, and that he can use his criterion of importance to be consistent about when he puts excellence ahead of fairness and fairness ahead of excellence. It is up to us to prove that he is wrong. *Perhaps* we can but if so, we will have to do it the hard way, after a full and fair debate using logic and empirical evidence as our principal weapons.

Thus far, I have used Nielsen's later works

²⁰ *Ibid.*, p. 91.

mainly to supplement the arguments used in his first major article, "The functions of moral discourse." However, some of these later works introduce new lines of argument and it is to these that I now turn. Here I will have to be selective. Two in particular appear relevant to our debate with elitists: when reasoning morally, some men build on a foundation of either scientific or metaphysical myths; others ignore the role of sympathy in moral reasoning.

Nielsen criticizes a variety of racists and sexists as assuming a mythical anthropology or sociology and, when imagining a Buddhist saying we should strive for nirvana, remarks that this makes sense only in terms of some "wildly metaphysical scheme."²¹ I have no desire to defend the concept of nirvana, but I had better say something more on behalf of my Social Darwinist. Given the long-standing historical connection between his ideals and racism, he may seem to have got off rather lightly.

I have heard a sophisticated Social Darwinist hold forth as follows: "In the long run, racism is an impediment to Social Darwinism—it posits an absurd criterion of human excellence as a competitor to the criterion we hold, it retards giving the strong the rewards and resources they merit and it saves many from going under who certainly ought to do so. However, historically there is no doubt we owe a great deal to racism, whether that of the common man or that of the late nineteenth century scholar who made metaphysical hash out of evolutionary biology; for these doctrines did much to condition men to accept the notions of a competitive social struggle and the superiority of the winners of such a struggle. But don't feel too superior. Many of your historians believe that the absurd metaphysics of Christianity performed a similar service for humanism, it conditioned men (however irrationally) to accept equalitarian ideals; and they still debate about whether humanism is strong enough to stand on its own feet. Racism has been an effective school for elitists—perhaps we are now strong enough to do without it or we may continue to need it for a substantial transitional period. At any rate, while I cannot speak for my followers, I am an empiricist and

not a metaphysician; and I prefer my anthropology scientific rather than mythical."

As for the argument from sympathy, Nielsen begins by granting that universalizability is a purely formal principle, that one can "with impeccable logic" calmly assert that men unlike oneself should go under. However, he thinks we have a right to ask such an opponent to go beyond universalizability and engage his *emotions*. We can ask him to vividly imagine the plight of his victims, to empathize with them, to identify with them, to imagine he was like them and that was *him* going under—and see if this does not awaken a feeling of sympathy and convince him he should take into account what his victims want. Nielsen believes that if our opponent is capable of understanding and appreciating the feelings and wants of others, our tactic will work; and argues that if he is not capable of this, he is simply not reasoning like a moral agent.²²

In my opinion, If we used our tactic, a Social Darwinist would answer as follows: "What you forget is the psychological distance between myself and someone who is weak, lazy, and sentimental. It is not that I am incapable of empathizing with him—it is just that I can not *both* empathize with him and remain myself! If I am to imagine myself going under, I have to fall short of identifying with him—that certainly does awaken outrage, but so what: my principles don't demand my sort going under. And when I truly identify with him 'I' disappear entirely and what I experience is *him* going under, *not me as him* (how could I be like him and still be like myself!) And thus empathy hardly leads to sympathy, recall he is the sort of man I loathe. However, just for the sake of argument, let us assume something new happens: that I am in fact overcome by sympathy. Does this show I *ought* to be? Does it not merely raise the question of whether he merits sympathy—and the question of whether it is a defect in me to feel it? Or assume something more: that when overcome by sympathy, I feel that such men ought not to go under. Do I not have the right to return to normal and make my moral assessment in a sober and calm state of mind? This is all conjecture, but if I did so, I believe I would

²¹ Nielsen: "Justification and Moral Reasoning," *op. cit.*, pp. 105-108; "On Moral Truth" in Nicholas Rescher (ed.), *Studies in Moral Philosophy* (Oxford, 1968), pp. 17-19 and 23-24.

²² Kai Nielsen, "The Good Reasons Approach Revisited," *Archiv für Rechts und Sozialphilosophie*, vol. 50 (1964), pp. 478-480.

say: 'how could I have been so weak as to be swayed by sympathy?—at any rate, I must not let it interfere with what I know to be right'."

In other words, the author believes that Nielsen's questions are in order. If a Social Darwinist is not willing to assess the consequences of his principles, we have a right to be suspicious; and to assess the consequences of his principles, he must know both the amount of suffering they entail and the sort of man they condemn—he might get to know the former in the dentist's chair (or through a series of personal misfortunes) but for the latter, he needs empathy. However, I also think that the Social Darwinist's reply is plausible and psychologically sound. As Katz points out, empathy involves a two-step process: when we actually identify with someone else, the "I" does disappear; and then, when we break off the relation of identity, the "I" re-emerges and we assess what we have experienced.²³ It is quite possible to have real empathy with someone and then, when the spell is broken, be disgusted by the personality we entered into—as every actor who has lived the part of Hitler on stage knows. Or take Charles Dickens, he certainly "understood" Uriah Heep, but he did not "appreciate" him, did not respect him. Empathy is feeling what others feel, sympathy is that plus having positive feelings towards others. The two are not the same and I am unwilling to allow Nielsen to make sympathy (for *all* men—an elitist may feel sympathy for his peers), much less being swayed by sympathy, a prerequisite for reasoning like a moral agent. In order to bridge the gap between them, Nielsen lays great stress on what the elitist is likely to feel when under the spell of empathy, after all, when he is actually identifying with his victim, won't he at that moment not want to go under? But I doubt that any elitist ever thought his victims wanted to go under, he just does not believe that what some men want counts for very much.

Well then, at this point in the debate at least, the realm of the moral approach has led to an unsatisfactory result. It has merely provoked a variety of accounts, humane, Nietzschean, Social Darwinist, aesthetic elitist, no doubt we could add others, all of which seem partisan and none of which seem to be uniquely viable. Nielsen himself may have concluded that his arguments are not decisive. In 1968, he published an extraordinary article in which he expresses scepticism about the notion of human rights—on the grounds that he doubts his ability to show that "a Nietzschean concept of morality" is mistaken! He goes so far as to say that a "Master Morality" qualifies as a moral point of view, thanks to being universalizable and embodying some kind of social concern (concern for at least some persons other than oneself).²⁴ If Nielsen were really to accept this last, the whole of the realm of the moral approach would begin to erode: the choice would no longer be between being Nietzschean and being moral. Nielsen would have to broaden his criteria for the realm of the moral and for every ideological opponent allowed to enter (with no strings attached), the approach would be just that much less useful in solving the problem of justification.

It is, therefore, hardly surprising that his scepticism article has attached a long and agonized footnote. And it seems to have had a long-range effect on Nielsen's confidence: beginning in 1971, articles citing the moral point of view with apparent confidence²⁵ alternate with those expressing grave doubts about our ability to answer the sceptic²⁶ and sometimes, these two themes are struck in the same work.²⁷ Most moving of all is a passage in which Nielsen reveals how much it means to him to refute the sceptic, he thinks of scepticism as a bulwark of bourgeois individualism and as destructive of fundamental criticism of our social ills.²⁸

* * *

²³ Robert L. Katz, *Empathy: Its Nature and Uses* (New York, 1963).

²⁴ Kai Nielsen, "Scepticism and Human Rights," *The Monist*, vol. 52 (1968), pp. 579–584—also pp. 592–593.

²⁵ Kai Nielsen, "When Are Immoralities Crimes?" *Philosophia*, vol. 1 (1971), pp. 135–136; and "The Enforcement of Morality and Future Generations," *Philosophia*, vol. 3 (1973), pp. 443–444.

²⁶ Kai Nielsen, "Anthropology and Ethics," *The Journal of Value Inquiry*, vol. 5 (1971), pp. 265–266; "Varieties of Ethical Subjectivism," *Danish Yearbook of Philosophy*, 1972, p. 87; and "On Locating the Challenge of Relativism," *Second Order* (1972), pp. 14 and 22–24.

²⁷ Kai Nielsen, "On the Choice Between Reform and Revolution" in Virginia Held, Kai Nielsen, and Charles Parsons (eds.), *Philosophy and Political Action* (New York, 1972)—compare p. 45 with p. 47. Also *Ethics Without God*, *op. cit.*—compare pp. 59 and 62–63 with p. 91.

²⁸ Kai Nielsen, "Is Empiricism An Ideology?" *Metaphilosophy*, vol. 3 (1972), pp. 269–273.

Thesis Three: Even if we can show that competing accounts are non-viable, the realm of the moral approach will be proved useless in the process.

In the last section, I tried to show that Nielsen's arguments do little to embarrass the Nietzschean or Social Darwinist and I think that this is worth showing. However, the fact that Nielsen's arguments have failed does not mean that further argumentation would fail, indeed, I myself believe that carrying the debate some steps further might pay dividends. And this raises a question: am I merely saying that the realm of the moral approach has not been vindicated thus far—and that better arguments might do so; or am I saying that there is a defect in the approach which allows us to reject it with some finality? I want to assert the latter and thus my third thesis.

We must take another look at the arguments Nielsen uses when debating against competing accounts of the realm of the moral. He tried to convict his opponents of misusing words, ignoring the facts of human nature (man is "the sort of animal to whom . . . moral blame and praise attach"), logical inconsistency, ignoring empirical evidence, bad metaphysics, and lack of empathy. What strikes one about these arguments is the overall pattern, the *familiarity* of the pattern; these are exactly the kind of arguments one might use when debating moral principles or, for that matter, aesthetics or even Freud versus Skinner. Which is to say that if Nielsen had won his debate about competing accounts of the realm of the moral, he could have used the very same arguments to win a debate about moral principles. Take the Nietzschean account, "using most of mankind simply as a means to the ends of superior men." If we can use the above arguments effectively against this assertion stated as an account of the function of ethics or as an account of the moral point of view, we could use them against the very same assertion stated as a moral principle. If it misuses language, or distorts human nature, or leads to logical inconsistencies, and so forth, certainly it is discredited in any event.

Just to spell all of this out: The realm of the moral approach is supposed to help us rout our ideological opponents. But to use it, we must argue for our own account or criterion of the realm of the moral. Let us imagine the optimum result; we argue down every one of our oppo-

nents and show that their accounts are non-viable. Now we can use the realm of the moral approach, but only because we have *already won* the debate against our opponents! We cannot use it when we need it (before we have won) and we can use it only when we no longer need it (after we have won). Moreover, in arguing about accounts, we have not come up with anything really new—we argue just as we would if we were arguing about moral principles. So we might just as well stick to arguing about moral principles and forget about "various accounts of the function of ethics" and "conflicting accounts of the moral point of view." In vindicating the realm of the moral approach, we have proved it useless in the process.

The assumption of this, of course, is that while Nielsen or Toulmin or Baier may do better using variations on the arguments listed above, they will not come up with any really new pattern of argument in the future. I cannot prove such a thing, I can only give my reasons for being pessimistic. When we argue for an account like "the function of ethics is to maximize happiness with equity," we are arguing for something just as *narrow* as a moral principle—or at least as narrow as our basic ideals or first principles. And I suspect that such a debate is very much determined by the content of what is being debated, same content, same opponents, same arguments likely to be effective against these opponents. It may be that a *broad* account can be defended by some very different pattern of argument, but that just will not do—it is not relevant to my point. At any rate, the burden is on our three thinkers to develop a new pattern of argument and show that it owes something to the approach they share. Until that time, I say that adopting something like "maximizing happiness with equity" as an account has only one effect: it transforms an argument which would normally be about first principles into an argument about competing accounts of the realm of the moral—and then we must cover exactly the same ground in terms of argumentation that we would have covered in the first place. So what is the point?

* * *

Thesis Four: If the realm of the moral approach is to be useful, it must be transformed.

In order to defend this thesis, I must present a list. Initially I am willing to call it a list that

delineates the realm of the moral; before we are finished it will have evolved into something else, a list of items likely to come up in moral debate. For the moment, however, let us say that in order to enter the realm of the moral, a man must:

- (1) Have ideals that take priority over his whims or desires—when “I ought” conflicts with “I want,” he must do the former or plead moral weakness.
- (2) Continue to hold his ideals when he gains a reasonably full knowledge of what they mean in practice—to do the latter, he will need empathy with other men (but not sympathy).
- (3) Be willing to universalize his ideals—be willing to give reasons for his moral assessments, at least until he reaches his first principles, and stand by his ideals with logical consistency throughout.
- (4) Have concern for the welfare of a group of men (the group may or may not include all mankind) rather than himself alone—this rules out pure egoism as distinct from invisible hand “egoism.”
- (5) It also rules out sadism which is a form of egoism.
- (6) He must possess some criterion of justice, at least justice as fairness, and he will probably have a criterion of justice as rewarding excellence as well—if so, he must have something to determine the balance between the two with logical consistency, e.g., a criterion of importance or significant life.
- (7) Use his ideals to pass judgment on his own (personal) way of life *in toto* as well as bit by bit.
- (8) Be willing to argue that his ideals, or a blueprint which aims as close to them as reality allows, are capable of ordering a human society.
- (9) Welcome the remaking of mankind in general in the light of his ideals—a certain diversity can of course be written into them.
- (10) Give happiness an important place—he need not give it top priority and certainly need not endorse the happiness of all mankind.

As the reader knows, the thinkers examined in this paper attempt to add items to this list, for example, an (11) about taking the interests of

every man into account, or a (12) about having to possess sympathy for mankind in general. And I have tried to drive them into a corner—they can add these items either by way of fiat or by arguing for them. I have said enough about the pointlessness of arguing for them, arguing for them as criteria of morality rather than being content to argue for them as moral principles. And if they add something like (12) by fiat, they merely invite an opponent to reply: “Ah well, if *that* is the only sense in which I fail to qualify as moral, you have won me over. Apparently, I pass everything else on your list so I can go forth with the message that my ideals are worthy of sacrifice, viable in the light of full knowledge, universalizable, alternatives to egoism and sadism, accord with justice, and provide men with a vision to order themselves, their society, and their species. All of this is enough for me—I have *no desire* to claim sympathy for all mankind! You will, however, excuse me if I draw a distinction between Morality I and Morality II and merely grant that I am non-moral in the first sense but not the second. You think you gain something by writing a humane criterion into your test for the realm of the moral, but all you do is get me to admit I am not humane—a price I am happy to pay.”

This reply shows why I have not worried too much about certain items on my own list. Items (9) and (10) are certainly subject to challenge, but I am willing to let anyone delete them who wants to pay the price they carry. If a non-ideological type feels ill-done by (9), and if he is a Calvinist as well and feels ill-done by (10), I will concede him the right to reply as above—“apparently the only sense in which I fail to qualify as moral is that I refuse to give happiness an important place and don’t care much about remaking mankind in general”—and I will concede him the same sort of distinction between Morality I and Morality II. Under these conditions, he should not resent the length of my list. If I were a realm of the moral thinker, that of course would be a different matter. Toulmin and Baier tend to treat this realm as if it were a nation-state. They want to define its boundaries, toss their opponents out, and strip them of all the privileges of citizenship. Nielsen is more complex. He sometimes prefers to convict his opponents of irrationality or factual error, rather than classify them as non-moral, but even he seems to think it terribly important to de-

termine where the "real" boundaries of the realm of the moral lie and use that as a weapon against his ideological opponents.

And now my secret is out. As far as the problem of justification is concerned, I do not care *where* we as philosophers, trying to delineate the realm of the moral, draw a line on my list, whether after (10) or (8) or even after (1). The *list itself* is all important. I care where an opponent draws a line, but only because of the items on the list and the prices attached to each. If he crosses (2) off his list, we could say something like "he recommends his ideals only because he averts his eyes from the suffering they cause." If he omits (3), we can say that he does not take his ideals seriously enough to stand by them with logical consistency. If he omits (4) or (5) or (6), we can say (among other things) that he refuses to give us any help in assessing what men either deserve or merit. And if he crosses off (7) or (8), we can ask him if his ideals are irrelevant to organizing his own life and to ordering a human society. These are the prices that count, not something called "lacking a moral point of view" or "not reasoning morally." If any of my ideological opponents wish to pay the prices attached to (1) through (8), all the better—it is just that I have difficulty finding any of historical importance who have chosen to do so.

I want to call attention to two advantages of our list: it will clarify *communication* with our opponents and it will clarify the problem of *justification*. As for communication, when someone uses terms like "moral" and "non-moral," a perfectly sensible thing to do, we can ask him how many of the items on our list he means to imply (he may of course want to add items as do our three thinkers) and then, we will have a precise notion of his usage. As for justification, where he draws his line on our list will usually tell us something of crucial importance—it will tell us what prices he means to contest and which he means to forfeit. And best of all, our list will keep us from being confused by an eccentric use of terms. Nietzsche for example did not like to call his ideals "moral"—and this has led many to assume (including the author at one time) that he would not try to justify himself in moral debate. But if we read Nietzsche carefully, and read him in the light of our list, we can see that he has no intention of letting any price on our list go by default. He was most interested in showing that he put his ideals

ahead of his desires (he loathed the hedonist), that he had faced up to their consequences, that he was being logically consistent about them, that he had a standard of "justice" (he focused mainly on excellence but fairness comes in for supermen), and that his ideals could be used to order a society (his new Europe). Nietzsche wanted to show these things because such prices can *stand on their own feet*—they are serious enough in themselves and if a man must pay them, it adds nothing to say "and besides that you are outside the realm of the moral." I sometimes think this last is what Nietzsche was getting at and if so, he was quite correct.

The core of truth in the realm of the moral approach is this: there is a structure to moral and political debate, there is an agenda men usually cover when they debate about their principles, and it is useful to have that agenda made explicit and to see why even bitter opponents, such as Nietzsche and a humanist, can agree to adopt it in common. After all, if we think about what actually happens, the parties to moral debate do not omit items from our list—rather they both *claim* that they need not pay the prices attached and then, each tries to show that *in fact* his opponent must. For example, the non-humane are likely to turn on us and accuse us of not really facing up to the consequences of our ideals ("it is your sort that allow haemophiliacs to live and reproduce"), and not really being logically consistent ("how do you square your admiration for genius with your equalitarianism"), and not really being able to order a human society ("you socialist-humanists would wreck any society foolish enough to let you rule"). The argument that ensues, about whose principles carry what prices, is the very soul of moral debate.

On the other hand, the realm of the moral approach must be transformed and now the reader can see what I mean by that. Our three thinkers make their mistake not in compiling a list but in how they tend to use it. They want to add items until they reach a certain boundary, a boundary narrow enough to exclude their opponents, and this last leads straight towards either an escape from real debate (as with Toulmin and Baier) or real debate about an artificial issue (as with Nielsen). We must give up the preoccupation with boundaries, give up arguing with Nietzsche about the function of ethics, and give up the notion that there is

something meaningful at stake called being non-moral. Rather we must use our list as an agenda, we must argue about ideals and principles and the many prices attached to the specific items on our list. It may be said "what if an opponent refuses to call it a list having anything to do with morals, or even denies its whole relevance to moral debate?" But we already know the answer to this—he can call it anything he likes

and ignore it at his peril—I will go on using it against him in any event.

It is not that Toulmin and Nielsen and Baier never argue as I recommend. If I am correct about the structure of moral and political debate, they could hardly help but do so. But they should argue only in that way and give up all that is distinctive about the realm of the moral approach.

University of Otago

Received August 28, 1975

IV. DIRECT AND INDIRECT REALISM

ANDREW WARD

THE causal processes of perception have traditionally been taken by upholders of what is now called indirect realism, and what used to be called representative realism, as one of the most serious objections to maintaining a direct realist theory of perception. Direct realists, in their turn, have traditionally pointed to the inferential nature of indirect realism as one of the most serious objections to that version of realism. For, as the name implies, according to an indirect realist no one can ever perceive external objects directly. An observer must rely, instead, on his ability to explain the order of his past perceptions, and predict the order of his present perceptions, as providing him with inferential evidence for the following claim: that *the best hypothesis* to account for the sequence of his perceptions is that these perceptions, though themselves only the awareness of mental images (sense data), have been caused by the existence of (directly unobservable) external objects. And the fact that an indirect realist has had to support his claim to know that external objects exist, by appealing to this inferential evidence, has traditionally been taken as a weak point in his position.

The aim of this paper is to examine a direct realist's defence of his theory in the face of the physical and physiological processes involved in perception. It is important to note that the existence of these causal processes is accepted by direct realists. What I shall try to show is that whereas these processes pose very serious problems for direct realists, they pose no corresponding problems for indirect realists. I shall also try to show that a direct realist has to rely on using a closely similar kind of "best hypothesis" argument for supporting his claim to perceive an external world directly, as an indirect realist relies on for supporting his claim to know that an external world exists, although never perceived directly.

I

The bare outlines of the causal processes occurring in visual perception are these: visual

perception only occurs if light quanta (photons) have been reflected or emitted from external objects onto the observer's optic nerves, and the electric impulses thus generated conveyed, through the afferent nerves, to the occipital lobes situated at the back of the brain. (Resembling causal accounts can be offered for the other senses.)

Now a direct realist claims that the existence of these causal processes should not lead us to think of the observer's resulting perceptions as the mere awareness of sense data. Far from viciously intruding between the observer and an external object, so blocking him from directly perceiving that object, the causal processes should be understood as producing the observer's direct perceptual act of awareness of the external object. That is, in cases of veridical perception, a direct realist maintains a version of what is called the act/object theory. The story is that when the occipital lobes are stimulated via the normal causal processes between object and lobes, this stimulation simply causes (or is identical with) the observer's direct perceptual *act of awareness* of the external object.

It is partly because he distinguishes, in the case of veridical perception, between the observer's *act of awareness* and *what* he is directly aware of (the external object) that the direct realist usually proceeds to argue that the causal processes are merely a *physical precondition* of direct perception, i.e., are merely processes which, contingently, must always occur if the observer is to be able to perceive objects directly. It is admitted that, *as a matter of fact*, an observer would not be able to perceive an external object unless the causal processes produced his awareness of the object. But it is not usually admitted that our grounds for deciding whether someone can be directly perceiving an external object need make any reference to the causal processes. Or to put the point in a more formal way, the causal processes of perception need not enter into the logical analysis of perception at all. Whereas a person can only make a logically

justified claim to be seeing the real color of a ripe tomato, if, for instance, he is not color-blind and the object is in sunlight, a person can make a logically justified claim to be perceiving an object directly, without cognizance being taken of the causal processes involved in his (claimed) direct perception. This is what is meant by saying that the causal processes are simply a *physical* precondition for directly perceiving an object. While "not being color-blind" is a *logical* precondition for determining whether a person is seeing the color of an object, the causal processes do not enter into the logical analysis of perception at all. They are simply processes which *contingently* help to explain an observer's ability to perceive an object directly. They are a *physical precondition* of an observer's direct perception because, it has been found, *without the occurrence* of these causal processes an observer will not, as a matter of fact, have the ability directly to perceive an object. But the logical criterion, for deciding whether an observer can be perceiving an object directly, makes no necessary reference to the operation of the causal processes between that object and the observer's occipital lobes.

Actually, it is untrue to say that the logical criterion for deciding whether an observer can be directly perceiving an external object need take no cognizance of the causal processes. If direct realists only had to contend with the scientific evidence so far discussed, it would indeed seem unnecessary to include any reference to the causal processes within the logical analysis of perception. But direct realists are forced to insert a causal clause because of a further scientific discovery. It has been found that, by artificially stimulating the occipital lobes with electric probes, an observer can be made aware of a perceptual field which is, or at least always appears to the observer to be, qualitatively indistinguishable from the perceptual field he would be aware of, were his awareness produced by an identical electric current reaching the occipital lobes via the normal causal intermediaries between object and lobes.

Like it or not, direct realists accept that the perceptual field an observer is aware of through artificial stimulation is a sense datum. But they do not accept that, *mutatis mutandis*, the perceptual field an observer is aware of, when his awareness is caused in the normal way, must equally be identified with a sense datum. In

short, direct realists do not think that the results obtained, through the artificial stimulation of the occipital lobes, force them to drop their theory in favour of indirect realism. They do not think they are forced to drop their theory because, they claim, we can distinguish cases of veridical direct perception, from cases of the mere awareness of a sense datum, by making it a *logical* precondition for the occurrence, on any particular occasion, of a veridical direct perception that the observer's awareness of a perceptual field, on that occasion, should be caused in the normal way. So, where an observer's awareness of a visual perceptual field is only artificially caused by electric probes, it would follow that the observer cannot be directly seeing the external object, which it appears to him that he is seeing, *even when* there is at the time that very external object before his eyes. On the other hand, granted that the observer's awareness of a perceptual field has been caused by light quanta coming from the external object and stimulating the occipital lobes through the normal causal intermediaries in his brain (and granted that any other logical preconditions have been satisfied), then it does follow that the observer can be seeing that external object directly. (Though not necessarily under every possible description.)

A direct realist, then, claims to circumvent the difficulty, raised by the possible artificial stimulation of the occipital lobes, by ceasing to maintain that the causal processes are simply a *physical* precondition for directly perceiving an external object. He now claims, instead, that the causal processes are a *logical* precondition: only if an observer's awareness of a perceptual field has been caused in the normal way can that observer be perceiving an external object directly. If the normal causal processes are not operating, all that an observer can be experiencing is a sense datum.

II

One comment ought immediately to be made about this defence of direct realism. It cannot reasonably be held to account for the results obtained through the artificial stimulation of the occipital lobes in anything like as plausible a fashion as the account given by indirect realists. Direct realists, however, claim that all the facts concerning the causal processes can be accommodated *at least* as plausibly within their theory as within the rival indirect theory. This

is, indeed, one of their chief grounds for championing a direct rather than an indirect realism. For, as I pointed out in my introductory remarks, it has traditionally been the alleged difficulty of squaring the existence of the causal processes with direct realism that has led indirect realists to adopt their theory. Yet if, as direct realists assert, the operation of the normal causal processes between object and occipital lobes is a *precondition*—whether physical or logical is not important here—for an observer directly to perceive that external object, absolutely no reason has been given for the following phenomenon. Why, when only *part* of this precondition is operating (i.e., when, by electric probes, *only* the occipital lobes are stimulated), can an observer be made aware of a perceptual field which is, or at least always appears to the observer to be, qualitatively indistinguishable from the perceptual field he would be aware of, were his awareness produced by the operation of the *full* precondition (i.e., when the observer's awareness of a perceptual field is caused in the normal way)?

There are really two points that are left entirely unexplained. First, it is left entirely unexplained *why*, when only *part* of a precondition is operating (i.e., when *only* the occipital lobes are operating) the observer should be aware of a perceptual field *at all*. Second, it is left entirely unexplained *why* what the observer does experience (a mere sense datum) should be, or at least always appear to the observer to be, *qualitatively indistinguishable* from the perceptual field he would experience, if the occipital lobes had been stimulated by an identical electric current, though stimulated in the normal way.

On the direct realist's account, it apparently just has to be accepted as a brute fact that when an observer's occipital lobes are stimulated via the normal causal intermediaries, the observer is *simply* caused to be *aware* of a perceptual field; while if only the final part of this causal chain is stimulated (i.e., if, by electric probes, only the occipital lobes are stimulated), what is caused is not *simply* the observer's *awareness* of a perceptual field but *also* a *perceptual field*. And, what is more, the artificially produced perceptual field—agreed by both direct and indirect realists to be a mere sense datum—will always appear to the observer to be *qualitatively indis-*

tinguishable from the perceptual field he would be aware of, were his awareness produced through an identical electric current stimulating the occipital lobes via the normal causal intermediaries. No explanation is offered for these remarkable occurrences—remarkable, within the context of a direct realist theory, that is.

An indirect realist, on the other hand, can perfectly plausibly explain the results obtained through artificial stimulation of the occipital lobes. For, of course, he maintains that even cases of veridical perception (i.e., cases where the awareness of the perceptual field is caused in the normal way), are cases of the awareness of a sense datum. Moreover, an indirect realist makes no distinction between the *act* of awareness and *what* one is aware of. He holds, instead, what is called an adverbial theory, according to which a sense datum never exists independently of one's awareness of it: the awareness-of-a-sense-datum is thought of as caused by (or identical with) the stimulation of the occipital lobes. So, on an indirect realist's account, there is evidently nothing surprising in the fact that an *artificial* stimulation of the occipital lobes can produce the awareness-of-a-sense-datum. Nor is it surprising that the *content* of this sense datum should appear to be (since for an indirect realist it actually is) qualitatively indistinguishable from the content of a perceptual field—always a sense datum, for the indirect realist—produced by an identical electric current stimulating the occipital lobes via the normal causal intermediaries. Indeed, these occurrences are just what one would have *predicted* on his theory:

(I shall not now discuss how an indirect realist *verifies* that his perception has been caused in the normal way. But it is clear that there is a parallel between his method of differentiating a veridical perception from a delusive one and a direct realist's. In order to differentiate a genuine direct perception from the mere awareness of a sense datum, a direct realist must, as we have seen, verify that the observer's awareness has been caused in the normal way. Likewise, in order for an indirect realist to differentiate a veridical [indirect] perception from a delusive one, he must verify that the observer's perception has been caused in the normal way.¹ As we shall see later, when I say something about the

¹ As it stands, the condition that a veridical *indirect* perception only occurs if the observer's awareness-of-a-sense-datum has been caused in the normal way is almost unquestionably too strong. I imagine it would be generally agreed

"best hypothesis" argument, the difficulties in verification encountered by an indirect realist have corresponding difficulties for a direct realist.)

III

There is an interesting parallel between the difficulties so far raised for direct realists and difficulties raised for them by another feature of the causal processes. This is the existence of a time interval between—in the case of visual perception—light quanta leaving an external object and the observer's ostensible direct perception of that object. Owing to the finite velocity of light, there is a time interval between photons leaving the external object and striking the optic nerves. There is a very slight further time interval between the stimulation of the optic nerves and the resultant electric currents reaching, via the afferent nerves, the occipital lobes. We have already noted that an observer does not perceive anything unless the occipital lobes are stimulated. (As one would expect, there is also a time interval, of a similarly small magnitude to that occurring between optic nerves and occipital lobes stimulation, before any tactual perceptions occur.)

Indirect realists have frequently maintained that the existence of the time interval provides a good reason for accepting their theory rather than direct realism. For, they have argued, how can one possibly claim to be directly perceiving an external object when the content of one's perceptual field, owing to the time interval, may be very different from the actual state of that object at the moment of one's perception? In reply, direct realists have pointed out that this argument makes use of a tacit assumption. It tacitly assumes that a direct perception can only occur if the *present* content of an observer's perceptual field is identical to the *present* state of

the external object. Direct realists have argued that no warrant has been given for this assumption. Why *should* the evidence of the time interval be taken to show that perceptions are the mere awareness-of-sense-data mediately caused by (numerically distinct) external objects? Why should not the evidence be taken to show that we have the ability *directly* to perceive external objects *in the past*? All the scientific evidence need be taken as showing is simply that we always directly perceive objects as they were and not, as we everyday assume, as they now are.

So far as I can see, there is nothing logically contradictory in the direct realists' rejoinder. But what does not seem to have occurred to them is that their reply leaves a number of facts quite unexplained. A number of facts which, on the indirect realists' theory, have perfectly predictable explanations. Here again, I must emphasize as strongly as I can that direct realists maintain that all the facts involved in the causal processes of perception can be *at least* as plausibly accounted for on their theory. My point is that this is simply not so: that indirect realists have no less clear and much more comprehensive explanations.

We may concede that, since the operation of the normal causal processes is a precondition of veridical perception, there is bound to be a time interval—corresponding to the time taken for photons to come from the external object and stimulate the occipital lobes—before that external object will mediately cause an observer to be aware of a visual perceptual field. But this concession leaves it quite unexplained why the *content* of the perceptual field should be that of the external object at some time past; and indeed at the *exact* time in the past when the photons, which are now causing the present stimulation of the occipital lobes, left the external object. If the causal processes only produce

that if, for instance, an observer's optic and afferent nerves were damaged, so that occipital lobes stimulation had to be effected artificially, this would not necessarily disqualify the observer from experiencing a veridical indirect perception. (Even a direct realist can arguably allow *certain* devices to replace normal perceptual organs, before conceding that the perception cannot be direct.) But for ease of exposition, I shall assume that the indirect realist's criterion, for the occurrence of a veridical indirect perception, lays down that the perception must be caused in the normal way. The essential point is that, however much one decides to relax this criterion, the indirect realist will still need to provide an argument to justify the claim to know that his perception, as of an external object, has been caused by that external object—*whatever* are allowed as intermediate causal factors. Equally, however much a direct realist can *reasonably* be permitted to relax the criterion for what is to count as "caused in the normal way" (because of, e.g., the possible replacement of a damaged afferent nerve by a mechanical substitute), it still remains true that he will be distinguishing a direct perception from the mere awareness of a sense datum, by making it a logical precondition, for the occurrence of a direct perception, that the observer's awareness should be caused in the normal way.

the observer's *act of awareness* of an external object, one would instinctively have expected that once the occipital lobes *had* been stimulated in the normal way, the observer would see the object as it is *now*, not as it *was* when the photons, at present causing the occipital lobes stimulation, left the object.

A direct realist might protest at this juncture that *of course* one sees the external object as it was when the present causally efficacious photons left the object. After all, it is these photons that are *causing the perception*. But on a direct realist's analysis, it does not obviously follow that *what* the observer is directly aware of, should be the state of the external object as it was when the photons, which are now causing his perception, left that object. On the contrary, *what* the observer is directly aware of is only *obviously* correlated with the state of the external object, as it was when the present causally efficacious photons left that object, if one holds an indirect realist's theory. For since an indirect realist uses the term "perception" partly to mean "awareness-of-a-sense-datum," it follows that when *he* says that a given collection of photons are now "causing the perception," he means that these photons are causing not *just* the observer's present *awareness* but also (understood adverbially) *what* he is directly aware of, namely the sense datum. So, on an indirect realist's theory, one would naturally expect the sense datum to be correlated with how the external object *was* when the photons, now causing the sense datum (as well as the observer's awareness of it), left that external object. (An indirect realist need not hold that the sense datum is an exact representative mental copy of the external object: only that there is always a formal correlation between the two.) But when a *direct* realist talks of photons "causing the perception," he means only that the photons are causing the observer's *act of awareness* of the (independently existing) external object. The photons are not supposed to be causing *what* one is directly aware of. Evidently not; since what one is directly aware of, according to a direct realist, is the (independently existing) external object. Consequently the direct realist has provided no explanation for why *what* one is directly aware of should always be the state of the external object contemporaneous with the time when the photons, now causing one's act of awareness, left that external object.

For this reason, Bertrand Russell's example of the star that explodes, between the time photons leave it and the observer's present resultant occipital lobes stimulation, can be used to make a substantial point against the direct realist. Russell himself used the example in an attempt to show that since the star no longer exists, one cannot rationally maintain that the observer is now directly seeing the star. Direct realists have scoffed at this argument, on the ground that it assumes the tacit and unwarranted premise already noted, viz., it assumes that one cannot be perceiving an object directly, if the content of one's *present* perceptual field differs from the *present* state of the external object. But the example of the exploding star can be used to make one of the very points I have been trying to make above, although in a more dramatic and picturesque way. If the causal processes are only a precondition of an observer's direct perceptual awareness of the star, why should one see the star as it was *before* it exploded? In fact, why does one see *anything at all*, granted that the causal processes only produce the observer's *act of awareness* of the external object, and that there is in this case, *ex hypothesi*, no object now existing to be aware of—since the star exploded some time ago?

Perhaps the mere fact that direct perception must always be, however slightly, of objects as they were, is no more extraordinary than the pre-scientific instinctive belief that (at least visual and tactual) perception is always of objects as they now are. That may be so; but the point is that direct realists leave two crucial facts entirely unexplained. First, assuming that it is just a prejudice of common sense to suppose that direct perception *must* be of objects as they now are and *cannot* be of objects as they were, no reason has been given for why one *does* always see objects as they were. It is no good a direct realist appealing to the scientific facts to explain this. For although, owing to the finite velocity of light, it follows that the farther away an object is, the longer will be the time before an observer is *able* to see that object directly, no reason has been for why, once one *is* in a position to see the object directly, *what* one directly sees should always be a *past* state of the object. Second, no reason has been given for why the state of the object that one does directly see should always be *just* as that object was when the present causally efficacious photons left it.

On a direct realist's account, to say that one cannot see an object before one's brain has been stimulated by photons from that object, does not entail that, once one's brain has been so stimulated, *what* one sees must equally be the object as it existed at *just* the moment when the present causally efficacious photons left it. For instance, once one is in a position directly to see the object, why does one not see it as it was at some period *between* the time that the photons left the object and the resultant occipital lobes stimulation? (With the sense of touch, the past state of objects, ostensibly directly felt, always corresponds precisely to the varying times taken for an electric current to reach the brain from the different parts of the body where objects may be felt.)

At the risk of laboring the point, I would like to stress again that an indirect realist has a perfectly straightforward explanation for the existence of the time interval. Both direct and indirect realists accept that a veridical perception only occurs after the occipital lobes have been stimulated in the normal way. And since, according to an indirect realist, this stimulation causes, or is identical with, the awareness-of-a-sensendum (the indirect realist's notion of a perception)—and *not* just, as the direct realist claims, the observer's *act of awareness*—one would naturally expect the state of the object, indirectly perceived, to be as that object *was* in the past: for photons take time to travel from the object and produce occipital lobes stimulation. Moreover, since the photons will take longer to reach the brain, the farther they have to travel, the particular past state of the object, indirectly perceived, would naturally be expected to correspond (as both parties agree that it actually does) to the *precise* time taken for photons to travel from the object to the optic nerves and for the resultant electric currents to stimulate the occipital lobes.

In an attempt to escape from the implausibilities of his treatment of the time interval, a direct realist might admit that he had been wrong to hold that the causal processes were simply what gave an observer the ability to perceive an object directly. What he should have said, he might assert, is not simply that the causal processes give an observer the ability directly to perceive an object, but that these processes *also* determine how far back in time, from the moment of an observer's occipital lobes stimulation, an object will be directly perceived

by that observer. Thus in visual perception, the idea would be that, when a brain stimulation had been set up by a particular collection of photons, this brain stimulation would put the observer in a position directly to see, and only to see, the object at a given distance back in time *from the moment of brain stimulation*. The particular character of the collection of photons, now stimulating the brain, would determine how great this time gap would be (presumably by stimulating the occipital lobes in a specific way).

But this idea does not fit the facts. Imagine two qualitatively identical and synchronized clocks placed so that light takes five minutes to reach an observer from the nearer clock and ten minutes from the farther clock. Now provided that both clocks are illuminated with identical lighting, the observer can calculate that five minutes after photons reach him from the nearer clock, qualitatively identical photons will reach him from the farther clock. Say that at 12:05 (his time) he looks up at the nearer clock and sees it standing at 12:00. So, according to the direct realist, the observer can tell that the kind of photons reaching him from the nearer clock at 12:05 (his time) will put him in a position directly to see any given object, from which *like* photons reach him, as that given object was *five minutes before the time that the photons reach him from that object*. Hence if, at 12:10 (his time), he looks up at the farther clock—at which time qualitatively identical photons will reach him from that farther clock as reached him from the nearer clock at 12:05 (his time)—he should directly see the farther clock as it was five minutes ago. That is, he should directly see the farther clock standing at 12:05. But, in fact, he will see the farther clock standing at 12:00. Therefore it is false that the stimulation of the observer's occipital lobes, by a particular collection of photons, puts the observer in a position directly to see an object at a given time interval back into the past from the moment of that stimulation. (A similar conclusion holds for touch.)

I had, earlier, implicitly guarded myself against the attempted escape just examined. For, in the original discussion, I pointed out that an object is always seen as it was at the *precise* time that the present causally efficacious photons left the object. It is not the case that the observer will see an object at some constant

time gap back into the past after brain stimulation, irrespective of the distance travelled by, and the velocity of, the causally efficacious photons.

IV

This completes my discussion of some of the difficulties raised for a direct realist by the causal processes of perception. We may now summarily compare these difficulties. In the case of the possible artificial stimulation of the occipital lobes, the direct realist left totally unexplained two facts: (a) Why, when only *part* of a precondition for perception is operating (i.e., when, through electric probes, only the occipital lobes are stimulated) an observer should be aware of a perceptual field *at all*; (b) Granted that an artificial stimulation of the occipital lobes *can* result in the awareness of a perceptual field, why the content of that artificially produced perceptual field should always seem to the observer to be qualitatively *indistinguishable* from the perceptual field he would be aware of, were his awareness produced by an identical electric current stimulating the occipital lobes via the normal causal intermediaries. In the case of the phenomenon of the time interval, the direct realist gave no satisfactory explanation of two facts: (c) Why *what* an observer perceives should always be a state of an object in the *past* never in the *present*: (d) Granted that one *does* always perceive a past state of an object, why the past state of the object perceived should always correspond *precisely* to the instant in the past when the present causally efficacious photons left that object.

Whereas the direct realist has given no coherent explanation for any of the occurrences listed under (a), (b), (c) and (d), an indirect realist has both a predictable and a straightforward explanation for each one of them.

But if, with respect to the occurrences just summarized, indirect realism is so much more plausible than direct realism, can any countervailing arguments be offered for maintaining a direct realist's position? Two principal ones may be given. First, it may be claimed that one cannot conceivably employ the causal processes of perception to undermine direct realism, because the knowledge of these processes has itself been acquired by assuming the *truth* of direct realism. If one casts doubt on direct realism, one is *pro*

tanto casting doubt on one's knowledge of the very processes which led one to doubt direct realism in the first place. So any attempt to undermine a direct realist's position by employing the causal processes of perception is bound to be self-defeating.

Against this, it might be said that one's knowledge of the causal processes does not *necessarily* have to assume the truth of direct realism. One might acquire this knowledge *indirectly*. Indeed, if indirect realism is true, it must *necessarily* be the case that one's knowledge of the causal processes—if it really amounts to knowledge—is indirect. This answer, however, though correct so far as it goes, obviously does not fully meet the objection. It would only meet it, if one could also show that one *did* have indirect perceptual evidence of the causal processes. Fortunately, we do not need to discuss that difficult matter here. All that is now required is to point out that the direct realist *himself* accepts the existence of the causal processes. Their existence, according to him, has been discovered by directly perceiving events in the external world. Hence, if direct realism is an adequate theory, it must be able satisfactorily to accommodate their discovery within the theory. So all that is needed, in order to show up the serious implausibility of direct realism, is to *assume* its truth (as I have done), and then to show that the theory leads to grave implausibilities. The logician's notion of a *reductio ad absurdum* argument is that of assuming a set of premises to be true in order to show that collectively they lead to self-contradiction. What I have attempted to show is that accepting direct realism, though admittedly not involving one in self-contradiction, does force one to leave a number of crucial occurrences without any satisfactory explanation. Indirect realism, on the other hand provides satisfactory explanations for all these occurrences.

The second countervailing argument, urged in favour of direct realism, points to the difficulty confronting an indirect realist in justifying the belief in an external world. Since, according to his theory, one can have no direct perceptual knowledge of the external world, the ground for believing in its existence must be based on the claim that the best hypothesis to explain the sequence of (the majority of) one's perceptions is that these perceptions, though themselves only the awareness-of-sense-data, have been caused by directly unobservable material objects. The

trouble with the claim is to see just how the inference to the existence of unobservable entities (material objects) really is the best hypothesis to explain the sequence of one's perceptions.

This drawback, however, contrary to what is customarily supposed, has its analogue for the direct realist. We can see this by recalling that, in order to distinguish a veridical direct perception from the mere awareness of a sense datum, a direct realist needed to incorporate the causal processes within the *logical* preconditions for the occurrence of a direct perception: an observer can only be said to perceive an external object directly, if his awareness of a perceptual field (as of that external object) has been caused in the normal way. But how is it to be determined that an observer's awareness of a perceptual field *has* been caused in the normal way? If it is a necessary condition for a veridical direct perception to occur that the observer's awareness of a perceptual field should be caused in the normal way, then unless at *some* stage one brings in indirect evidence, it can never be verified that an observer's awareness *has* been caused in the normal way. For before any ostensible direct verification by the senses can be justified, it must be determined that the verifier's awareness of a perceptual field, ostensibly directly verifying that the normal causal processes have occurred, has *itself* been caused in the normal way. (Since it is a necessary condition for *any* direct perception to have occurred that it should be caused in the normal way.) And because the only method of *direct* verification of occurrences in the external world, and so of the causal processes of perception, is by the senses, it follows that without an appeal at *some* stage to indirect evidence, it can never be known that the normal causal processes are operating. Therefore it can never be known that a direct perception has occurred.

Accordingly, if a direct realist is to maintain his theory, he must put that theory forward as an inferential hypothesis which claims to be the best for explaining the sequence of his perceptions. The basis of his indirect proof will be very similar to that employed by an indirect realist to

justify his belief that his awareness-of-a-sense-datum has been caused by an external object. With a direct realist, the claim will be that the best hypothesis to account for the content of his present perceptual field, as well as the content of (the majority of) his past perceptual fields, is the existence of external objects moving about in regular ways (the laws of nature), and that these external objects are now being, as in the past they have been, *directly* revealed in perception. (It is, of course, only the last conjunct of this claim that is disputed by indirect realists.) Having thus shown himself (so he believes) to be justified in his claim now to be directly perceiving the external world, the direct realist can point out that it must, therefore, be the case that his present perceptual field has been caused in the normal way.

The situation then is this. Both direct and indirect realists share the common claim that the best hypothesis to account for the sequence of one's perceptions is the existence of external objects. Direct realists think of external objects as directly perceived, the objects themselves only causing one's *awareness* of them. Indirect realists think of external objects as perceived via the awareness of sense data, the objects themselves causing one's awareness-of-sense-data.

Whether the existence of external objects is the best hypothesis to account for the sequence of one's perceptions is clearly not something that I have here discussed. Though I am bound to say that I think it a claim which it is going to be very hard indeed to make out.² What I have tried to show is that both indirect *and* direct realists need to make it out. But should they succeed, it would not follow that both their theories are equally defensible. For both theories *also* need to explain how the causal processes of perception can be accommodated within their rival accounts of how one perceives the external world. At crucial places, direct realists provide no satisfactory explanations. Indirect realists do. To reverse a famous verdict of Hume's: direct realism contains all the difficulties of indirect realism, with some others that are peculiar to itself.

University of York

Received November 10, 1975

² I have attempted elsewhere to outline some of the difficulties about the "best hypothesis argument." See Andrew Ward, "What's Not Really Wrong With Phenomenalism," *Australasian Journal of Philosophy*, vol. 51 (1973), pp. 245-252.

V. PARASITICAL REFERENCE AND PARADOX

DOUGLAS ODEGARD

MANY referential expressions are used effectively only if they can draw nourishment from other sources. If I say "He forgot it," then there must be alternative means available for picking out what I intend my pronouns to pick out. My references are in this sense parasitical. If there is nothing to nourish them, then the result is a distinctive type of "nonsense." I do still say something, but what I say has absolutely no truth value, including the values "true," "false," "not true," "not false," and so on. My remark is in a sense *not true*, but not in any sense which underwrites the negative operation "It is not the case that he forgot it." My remark is not true in the way that a question is not true (or false, or nontrue, or . . .) and not in the way that a false statement is not true. Call such a remark, and the sentence used to make it, "degenerate."

A degenerate remark can be called a "statement" without being intolerably misleading. The sentence used to make it is syntactically correct and does not contain gibberish. There is a correct answer to "What did he say?", albeit not one which allows the remark's author to explain his parasitical reference. Of course, a degenerate remark is not a statement in any sense in which "statement" deliberately has the possession of truth value written into its use. But not every preanalytic sense of "statement" is of this sort. Although a degenerate remark cannot be a straightforward logical relatum because it lacks truth value, it can have derivative logical connections. For example, a degenerate "He took twenty more" cannot straightforwardly entail anything. Yet we can say that if the remark were true, then a statement of the form "He took at least ten more" would have to be true, that the former "entails" the latter in this sense.

Since degenerate statements are entirely without truth value, they should be distinguished from statements which incorporate category mistakes, which involve false presuppositions, which assign predicates to intermediate cases, or which have undecidable truth values. Statements of the last type do have a truth value, but one which is in some way beyond us; as in "Some number never considered specifically by man is such that were it so con-

sidered, it would be considered by exactly one million men." The other statements are sometimes said to "lack truth value," but in a way which underwrites a negative operation on the given statement. In saying that a category-mixing *S* fails to be true, we thereby endorse "It is not the case that *S*" and to that extent assign it a truth value. These are arguable points, of course. But in the case of degenerate statements, there is no room for a parallel argument. The lack of nourishment for the parasitical reference infects both the statement and any operation thereon.

Statements which refer to degenerate statements need not themselves be degenerate. If you assert a degenerate "He is eating something poisonous" because your parasitical "something" is unnourishable, then I can *truly* say "You said something degenerate," provided my parasitical "you" can feed on "the present reader of this paper" and my "something" can feed on "that he is eating something poisonous." If the provisos are unsatisfied, then my remark is just as degenerate as yours.

Self-referring statements at least frequently involve degeneracy. For instance,

(1) This statement is stated in English

involves a parasitical reference if it implicitly refers to what the author of (1) is stating in (1). Such a parasite is unnourishable. The question "What are you referring to, i.e. what are you stating in (1)?" cannot be answered by the author except by saying "that my statement in (1) is stated in English." This says no more than "that what I am stating in (1) is stated in English" and merely reinvents the same question. Understood in this way, then, (1) is degenerate.

Statement (1) can be contrasted with my other-referring statement that (1) is stated in English. If someone asks me "What are you referring to, i.e. what is the author of (1) stating?", I can reply "that his statement (1) is stated in English, i.e. that what he is stating in (1) is stated in English." I do not thereby fall back on a reference to what I am stating and hence I do not reinvent the same question. Therefore my statement about (1) is true and not degenerate.

Statement (1) also differs from

- (2) This is an English sentence.

The use of "this" in (2) is not unnourishably parasitical, since it does not involve an implicit reference to what is being stated in (2). By contrast, the use of "this sentence" in

- (3) This sentence is true

is unnourishably parasitical, since (3) is equivalent to "What this sentence is used to assert is in fact the case." I shall return to statements like (3) presently.

It may be that every purely self-referring statement is degenerate. If so, then the qualification "purely" makes the point relatively harmless. It allows us to assign truth values to

- (4) This statement, viz. that some statements are degenerate, is true.

It accommodates the point that if "this statement" in

- (5) This statement is made after 1974

is used to refer only to the making of (5), then (5) is not degenerate. And it lets us concede that not all self-referring sentences are degenerate. But the merit of the general thesis is not really important here. The important thing is to understand the degeneracy of those self-referring statements which are degenerate, in an effort to see through a number of ostensible logical paradoxes.

A logical paradox arises if there is a statement which has a given truth value if and only if it also has a logically opposing truth value; i.e. if there is a *contradiction* in this strong sense. Assuming that the given statement is genuine, the only way to resolve the paradox is to block the truth value assignments altogether. The notion of degeneracy is a potentially useful device for doing this.

Putative liar-type contradictions all involve unnourishably parasitical references. If I say "This statement is false," then I imply that what the negation of my statement asserts to be so is so. I thereby implicitly refer to what the negation of my statement asserts. Yet the only answer I can give to "What does your statement's negation assert?" is "that my statement is not false." This amounts to saying "that what my statement's negation asserts is not so," which simply reinvents the same question. My use of "this statement" is therefore unnourishably parasitical and my statement is degenerate. Similar observations hold for "This statement is not true," "It is not the case that this statement is true," and so on.

The statement "This statement is degenerate" is also degenerate. It amounts to saying "It is not the case that what is asserted either by this statement or by a negative operation thereon is the case." And there is no nonregenerative response available for "What does that statement or any negative operation thereon assert?". The same point applies to "This statement is either not true or degenerate."

Indirect self-reference can be treated in basically the same way. If I say

- (6) The other conjunct in (6) is true, and the other conjunct in (6) is not true

then both instances of "the other conjunct in (6)" are unnourishable parasites. The first conjunct in effect says "What my second conjunct asserts is so" and the only response to "What does your second conjunct assert?" is "that what my first conjunct asserts is not so." This prompts "What does your first conjunct assert?", to which the only answer is "that what my second conjunct asserts is so," thereby regenerating the original question. The same thing happens if we begin by querying the reference in the second conjunct. There is no way to break out of an explanatory circle. It is like engaging in the dialogue "He passed it" "What did he pass?" "What she received" "What did she receive?" "What he passed to her."

The same principles govern

- (7) The second conjunct in (7) is degenerate, and the first conjunct in (7) is degenerate.

Both conjuncts in (7) are degenerate and hence neither is true. If I try to explain the reference of "the second conjunct," all I can say is that my second conjunct asserts that it is not the case that what is asserted by my first conjunct or any operation thereon is the case. In trying to explain what I am talking about, I am therefore forced to appeal to what I am asserting, which is absurd.

The first conjunct in (7) is unlike my *subsequently* saying

- (8) The second conjunct in (7) is degenerate.

In (8) the reference to the second conjunct in (7) is still parasitical, but now it can be nourished. I can explain that the second conjunct in (7) asserts that it is not the case that what is asserted by the first conjunct in (7) or any operation thereon is the case. Since my reference to what the first conjunct in (7) asserts does not appeal to what I assert in (8), the circle is broken. Thus (8) is not degenerate but true.

J. L. Mackie is reluctant to go this far.¹ He describes a case in which the sentences " $2+2=5$ " and "There is no true sentence on this sheet" are the only things written on a sheet of paper. He agrees that we cannot say that the second sentence is true but wonders whether we might not contradict ourselves if we say that it is not true. In his words,

"Again we cannot say that the second sentence would make a true statement, and yet if we say . . . that it would not, then, since the only other sentence on the sheet would make a false statement, we are committed to using or endorsing a sentence which is another token of the same type and with the same reference as this one. How can we avoid contradicting ourselves here?"

His own conclusion is that we should decline to say, one way or the other. This is excessively timid.

Suppose I wave the sheet in the air and simultaneously write on the blackboard "There is no true sentence on this sheet." I do thereby refer to the same two sentences and my sentence is of the same type as the second sentence on the sheet. Also, my sentence and the second sentence on the sheet do in some sense have the same meaning. But it does not follow that they are used to make the same statements, if "are the same statements" entails "have the same truth value status."

In using the second sentence on the sheet, one implies that nothing asserted by a sentence on the sheet is the case. This implicit reference to assertions is parasitical and invites the question "What assertions by those sentences are you talking about?". A complete answer must include the reply "the assertion by my second sentence that nothing asserted by those sentences is the case." But one cannot explain a reference made while using a sentence simply by referring to the assertion made by that sentence. And the reference to what is asserted by the second sentence, viz. that nothing asserted by the sentences is the case, simply reinvents the original question. Hence the parasitical reference is unnourishable and the statement made by the second sentence on the sheet is degenerate.

The blackboard sentence, on the other hand, also involves a parasitical reference, but it can be nourished. When I use the sentence, I again imply that nothing asserted by a sentence on the sheet is so. But now if someone asks "What assertions are

you referring to?", I can answer partly with "the assertion by the second sentence on the sheet that nothing asserted by a sheet-sentence is so" and partly with "the assertion that $2+2=5$." I need not try to explain my reference partly by referring to the assertion I make when I use the sentence. I refer to the assertion made by the second sentence on the sheet. This allows me to break out of the circle and my statement is not degenerate. It is true.

Mackie's unqualified remark about the "same reference" of the on-sheet and off-sheet sentences can obscure things. The two sentences do refer to the same sentences and they might in a sense be said to refer to the same assertions. But their referential character in the latter respect is different. The parasitical reference of the on-sheet sentence cannot be nourished. Indeed, for certain purposes it might even be described as a failure of reference, or as a breakdown in the sentence's referential apparatus. In a sense, someone who says "It is hot" and cannot explain the reference of "it" does fail to refer. There is no such breakdown in the case of the off-sheet sentence. Hence the off-sheet sentence can be true while its on-sheet counterpart is degenerate. Mackie also obscures this point in other contexts when he says that "nothing is asserted" by "What I am now saying is false," and that such an utterance "lacks content" and raises "no substantial issue" (pp. 244, 259-262). If the defect in the on-sheet sentence were of this rather vaguely specified sort, then its off-sheet counterpart would appear to suffer the same defect. But the defect is not of this sort. It is referential. Hence the two sentences can be treated differently.

Before moving on to other putative contradictions, consider three solutions to liar-type paradoxes which are in some respects similar to, and yet importantly different from, the solution proposed here. Robert Martin argues that liar-type sentences are such that we cannot determine whether they commit category mistakes; hence we cannot assign truth values to them. He argues, roughly, that a subject-predicate sentence is free of category mistakes if and only if its subject's referent falls within its predicate's range of application. He claims that the range of application of "true" and "false" consists in sentences which avoid category mistakes. He then argues that it is impossible to determine whether a liar-sentence like "This sentence is false" is free of category mis-

¹ See J. L. Mackie, *Truth, Probability and Paradox* (Oxford, 1973), pp. 294-295.

takes, since doing so requires already having done so.³

Martin's proposal at most establishes that we cannot determine whether the problematic sentences are both true and not true. It does not show that they are not. Moreover, it draws on the contentious premise that assigning a negative truth value to a category-mixing sentence itself mixes categories, e.g. that we cannot without absurdity say "It is not the case that red is twice four, since red is not a number." Also, the proposal fails to establish the relevant defectiveness of sentences like "What this sentence asserts is interesting," "What I am now saying is carefully considered" and "This sentence is free of category mistakes." Finally, although Martin does discredit "This sentence either is false or mixes categories" by limiting a disjunctive predicate's range to that of its most restricted disjunct, he concedes that the limitation does not hold when the disjunction exhausts a category, as in "either true or not true." Consequently, he must endorse "This sentence either is true or is false or mixes categories."

Brian Skyrms holds that simple liar sentences like "This sentence is false" are "meaningless" in the sense that they are both not true and not false. He then seeks to avoid paradox at the higher level of "The sentence 'This sentence in double quotes is not true' is not true." He then claims that whereas a simple self-referring sentence is "caught in a semantical trap which renders it incapable of making a meaningful assertion," the higher-level sentence is not. He therefore refuses to move from the truth of the higher-level sentence to the truth (and hence also the non-truth) of the lower-level sentence. Indeed, he thinks that in general, if " $\sim Ta$ " ("Sentence a is not true," where ' a ' is the name of the sentence containing it) is meaningless, then $\sim T\sim Ta$ does not yield $\sim Ta$ by the substitution of a for " $\sim Ta$," even though $a = \sim Ta$.³

In formalizing the higher-level sentence as $\sim T\sim Ta$, Skyrms implies that it does assign a truth value to the lower-level sentence. In that case it is just as degenerate as the lower-level sentence and cannot be true. Moreover, dropping the substitutivity of identicals in truth value assignment contexts is an independently implausible move. On the other hand, if the higher-level sentence does not assign a truth value, then the symbolism should

be dropped and the account of "meaninglessness" sharpened. The claim that a sentence entirely lacks truth value avoids endorsing any negative operation on that sentence, and it is much more precise than simply saying that the sentence is both not false and not true.

Bas van Fraassen assimilates the logic of liar sentences to that of sentences with non-true Strawsonian presuppositions. He thinks that both "This sentence is false" and its negation presuppose the faulty "It is true that the given sentence is false." He recognizes that a further contradiction-candidate can always be formed by putting the negative value which results from a rejected presupposition in the blank in "This sentence is. . ."; e.g. from the non-truth of "This sentence is false" we get the candidate "This sentence is not true." He also admits that there is no expressible limit to such a series but finds the cost tolerable.⁴

The cost is more substantial than van Fraassen acknowledges. Suppose we let positive integers stand for permissible negative truth values: 1 = false, 2 = neither true nor false, 3 = neither true nor 2, etc. For any integer n , we can have a sentence which lacks all n values, viz. "This sentence is n ." But more seriously, we can have the transcendent sentence "This sentence has an integer value." If we try to assign it a negative value because of a faulty presupposition, it seems that we must assign it an integer value, which generates a paradox. Now, we might try to sidestep this result by claiming that the assessment of the transcendent sentence is not an integer evaluation and let a different numerical series represent such assessments. There would then be an infinity of infinite series. Yet as long as we are working with the model of faulty presuppositions, such a move is implausible. The model dictates that resulting assessments are truth value assignments belonging to a homogeneous series. Alternatively, if the model is abandoned or modified, then much of what is distinctive in van Fraassen's proposal disappears.

Putative heterological-type contradictions can also be dismissed as degenerate statements. If "heterological" means "not a property of itself" and I say " x is heterological," then I implicitly refer to the further thing that x is not insofar as x is not a self-property. My reference is parasitical and hence needs nourishment. Nourishment is avail-

³ See R. L. Martin, "Toward A Solution to the Liar Paradox," *The Philosophical Review*, vol. 76 (1967), pp. 279-311, and "A Category Solution to the Liar" in R. L. Martin (ed.), *The Paradox of the Liar* (New Haven, 1970).

³ See Brian Skyrms, "Notes on Quantification and Self-Reference" in *The Paradox of the Liar*, *op. cit.*

⁴ See Bas van Fraassen, "Truth and Paradoxical Consequences" in R. L. Martin (ed.), *The Paradox of the Liar*, *op. cit.*

able if x = being wise but not if x = heterologicality. If I say "Heterologicality is heterological," then the question "What is the further thing which heterologicality is not insofar as it is not a self-property?" prompts the response "heterologicality," which does not answer the question. The question is looking for a *further* property, to which the statement implicitly refers, and there is in principle no answer to this question. Consequently, the parasitical reference is unenrichable and the statement is degenerate.

It is important to note that the parasitical reference is to something *further*. Gilbert Ryle, who applies "heterological" to words, claims that in the case of "'Heterological' is heterological," "no property is ever mentioned, so the seeming reference to such a property is spurious."⁵ But the statement does in a way mention a property, viz. heterologicality. Similarly, he says: "to say that a word is heterological . . . is to refer to, without mentioning, some philological property (*not* heterologicality . . .) for which that word stands and which . . . does not belong to that word." But again, a critic might insist that heterologicality is just as "philological" a property as any other, and that heterologicality is referred to. To preserve the position against such criticisms, we should be clear that, in addition to the mentioning of the word "heterological" and the reference to the property heterologicality, there is also an implicit reference to what there is *further* about the word or the property in virtue of which it is not self-applicable. It is this reference which is both parasitical and unenrichable. And it is because of this that the statement is degenerate.

Similar observations hold for "'Autological' is autological." There is a parasitical reference to something further about the word in virtue of which it is self-applicable and when the word is "autological," then this reference is unenrichable. J. F. Thomson complains that this type of position fails to explain "why we should get a contradiction in the one case and not in the other."⁶ But the point is that we don't get a contradiction in *either* case, since both statements are degenerate. As Mackie points out, discrediting both statements is a virtue rather than a vice (p. 262).

Russell's ostensible class-membership paradox also involves a degenerate statement. As in the case of " x is heterological," the statement " x is not a member of itself" implicitly and parasitically re-

fers to something further which x is not, in virtue of which it is not a member of itself; e.g. the class of trees is not a self-member because it is not a tree. In the case of "The class of all classes which are not self-members is not a self-member," such a reference cannot be enriched. The question "What is that further thing which the class is not in virtue of which it is not a self-member?" cannot be answered by "a self-member" and yet this is all that one can say. Hence the statement is degenerate. A similar point holds for "The class of all self-membered classes is a self-member."

Adopting this solution for Russell's paradox allows us to retain the logical respectability of the property "not being a self-member" and of the class of all non-self-membered classes. We can say of such a class, e.g., that the class of trees is a member, and that the class of classes with at least two members is not a member. Moreover, we can endorse the correlation of classes and properties without fear of underwriting contradictions. Some philosophers will be usually suspicious at this point, however. I shall try to show that their suspicions lack independent foundations.

Mackie offers a rationale for refusing to endorse the class of all non-self-membered classes (see pp. 266-267). He agrees that there are non-self-membered classes but claims that there is no "determinate" class thereof because the property "not being a self-member" itself fails to be "determinate." He compares such a property with "being beside this," where the reference of "this" is not nailed down. He rightly claims that there can be no class defined on "being beside this" but wrongly suggests that "being a non-self-member" meets the same fate. What Mackie seems to be doing here is transferring the defect which characterises *ascribing* the property to the class defined on it to the property and the class. If we say that the class of all non-self-membered classes is not a self-member, then we do implicitly refer to something further about the class in virtue of which it is not a self-member. And this reference is defective. Hence the ascription is degenerate. Yet this reflects badly only on the ascription, not on the property or class. An "indeterminacy" in the ascription therefore does not generate a similar "indeterminacy" in the property or class. Granted, the property is such that ascribing it to a class does involve the indicated implicit and parasitical reference. But there is nothing intrinsically faulty about this. One

⁵ See Gilbert Ryle, "Heterologicality," *Analysis*, vol. 11 (1950-51), pp. 61-69.

⁶ See J. F. Thomson, "On Some Paradoxes" in R. J. Butler (ed.), *Analytical Philosophy*, 1st ser. (Oxford, 1962), p. 113.

should simply take suitable precautions when ascribing the property or talking about the class. Since the property can be assigned to classes without degeneracy, therefore a class can be defined on it. Indeed, once the reference of "this" in "being beside this" is nourished, say by adding "i.e. adjoining my ash tray," then a class can be defined on it as well.

Some philosophers draw a close analogy between the class of all non-self-membered classes and the barber who shaves all and only non-self-shaving barbers. The barber cannot without qualification exist, since his being a self-shaver mutually entails his not being a self-shaver. If there is "such" a barber, then he is bearded, or has a hairless chin, or in some other way falls outside the set of barbers in relation to which he is defined (e.g. they are French whereas he is German). In the same way, according to some philosophers, there is a class of non-self-membered classes only if it somehow falls outside the class of classes in relation to which it is defined. There cannot be "such" a class without some such restriction.

W. V. Quine thinks the barber/class analogy is logically exact, that the only difference consists in our current uncritical habits of thought.⁷ Thomson agrees on the exactness of the analogy, but feels that the motives for seeing a disanalogy are subtle and not easily removed. He offers the following logical theorem which he thinks both the barber and the unrestricted class of all non-self-membered classes violate (pp. 104-105).

Let S be any class and R any relation defined at least on S . No member of S has R to all and only those members of S which do not have R to themselves.

He then points to two plausible assumptions which he thinks can lead us to endorse the class of all non-self-membered classes in spite of the theorem (pp. 117-118). First, he thinks that we readily assume that such a class can embrace certain members of a given class of classes and also be a member of that class, i.e. that there is a class of all classes which satisfies this requirement. He thinks that we can just as readily drop this assumption: "The shortest and simplest way with the paradox is simply to deny that there are or could be two sets . . . related in this way." Second, he thinks that we often overlook a logical error in "For all classes x , x is a member of the class of all non-self-membered classes (C) if and only if x is not a member of x "

because we naturally assume that C is not itself a value of x . The assumption is for him correct. Where we go wrong is in also assuming that there is no inconsistency between the first assumption and accepting the class of all non-self-membered classes. Once we spot the inconsistency, we should reject the class.

Thomson is supposed to be offering an *independent* diagnosis of the mistakes which lead us into paradox. Consequently, he should not cite the avoidance of paradox as a reason for abandoning a given assumption. Yet this is precisely what he does in his first diagnosis. Aside from its offering a "way" to deal with paradox, no reason is given for dropping the indicated assumption. As a result the assumption's initial plausibility is left untouched. For instance, among the members of the class of all classes with at least three members we can locate both the class of all classes with at least two members and all its member classes. It *seems* that we can find a sufficiently expansive class of classes to do the same sort of thing with the class of all non-self-membered classes. Nothing in Thomson's diagnosis shows that this impression is illusory.

The alleged logical error in the second diagnosis is genuine only if C 's being a value of x obliges us to affirm or deny " C is a member of C ." But there is no such obligation, since we can accept C as a value of x and at the same time maintain that " C is a member of C " is degenerate. Consequently, we do not need the assumption that C is not a value of x in order to keep us from detecting an error. Therefore we do not have to rely on an assumption which is inconsistent with endorsing the class of all non-self-membered classes.

Thomson's logical theorem is sound if it says that no member of S has R to all and only those members of S which do not have R to themselves, *provided* the insertion of R between any two members of S has a truth value and provided the two members can be identical. He himself adds the two provisos when he lets " R is defined on S " mean "For any (possibly identical) x and y in S , saying that x has R to y has a clear true or false sense." But contrary to what he suggests, endorsing the class of all non-self-membered classes need not violate the theorem when understood in this way. We can endorse the class and yet condemn "The class is R to itself" as degenerate. In so doing we do not maintain that the insertion of R between any two possibly identical members of S has a truth value.

⁷ See W. V. O. Quine, "The Ways of Paradox" in *The Ways of Paradox and Other Essays* (New York, 1966).

Hence we do not violate the theorem. On the other hand, if the theorem is recast without the provisos, then there is a good chance of a violation. The class of all non-self-membered classes is a class. Hence it is a member of the class of classes (S), which has as members (is related by R to) all and only those classes (members of S) which are not self-members (are not R to themselves). But if the theorem is suitably recast, then it will no longer be the theorem which emerges from a rejection of the barber. Consequently, the close analogy between the class of all non-self-membered classes and the barber is still eliminated. Moreover, there is no independent reason to think that it will be a plausible theorem.

Solving paradox by appealing to degeneracy has the following advantages: It works for a broad range of putative contradictions and for their affirmative counterparts. It does not deny that putative contradictions are statements, or that they

say or assert something, or that they have content. It does not assimilate putative contradictions to category mistakes or to statements with faulty pre-suppositions. It can retain the class of all non-self-membered classes and can straightforwardly correlate classes with properties. It plugs into a possible theory of reference, a theory which can be tested by applying it more extensively to independent cases. It does not itself require dropping standard logical laws for truth-valued statements, although it does hold that statements can be entirely without truth value. There are important questions which remain unanswered, of course: notably, whether degeneracy can be utilised when considering Berry's and Richard's paradoxes, and whether a notation can be developed to exclude degenerate statements from the scope of a formal language. But as far as I can see, these are further considerations, not ones which are essential to assessing the position outlined here.

University of Guelph

Received November 7, 1975

A new journal

Linguistics and Philosophy

An International Journal

Managing Editor: ROBERT WALL (University of Texas, Austin, Tex., U.S.A.)

Executive Editorial Committee: Max Cresswell, Østen Dahl, Hans Kamp, Lauri Karttunen, Asa Kasher, P. Stanley Peters, Jr.

Editorial Board: John Anderson, Lennart Aqvist, Emmon Bach, Eve V. Clark, David Dowty, Michael Dummett, Charles Fillmore, Janet Fodor, Gilbert Harman, Jaakko Hintikka, Steven Isard, S. Y. Kuroda, Nicolas Ruwet, Jerrold Sadock, John Searle, Petr Sgall, Robert Stalnaker, Arnold Zwicky, William C. Watt.

Linguistics and Philosophy is a journal for studies focused on natural language which is of interest to both disciplines mentioned in the title. Although the area thus described is so extensive that a complete listing of relevant topics is precluded, at least the following specifically fall within it:

- traditional areas in the philosophy of language such as meaning and truth, reference, description, entailment, speech acts
- traditional areas of linguistics such as syntax, semantics and pragmatics (when the studies are of sufficient explicitness and generality to be also of philosophical interest)
- systems of logic with strong connections to natural language: modal logic, tense logic, epistemic logic, intentional logic
- philosophical questions raised by linguistics as a science: linguistic methodology, the status of linguistic theories, the nature of linguistic universals
- philosophically interesting problems at the intersection of linguistics and other disciplines: language acquisition, language and perception, language as a social convention

Two volumes will be published yearly. Subscription price per volume of 3 issues (approx. 450 pages) Dfl. 90,- / US \$36.00, plus Dfl. 12,50 / US \$5.00 for postage and handling.

Private persons (no institutions, libraries, etc.) may subscribe at the reduced rate of Dfl. 42,50 / US \$17.00 per volume, including postage and handling.

Subscriptions should be sent to D. Reidel Publishing Company, P.O. Box 17, Dordrecht-Holland, or to any subscription agent.

Private subscriptions should be sent direct to the publisher.



D. Reidel Publishing Company

P.O. Box 17, Dordrecht-Holland

Lincoln Building, 160 Old Derby Street, Hingham, Mass. 02043 - U.S.A.

VI. WHAT YOU DON'T KNOW WON'T HURT YOU?

JOHN A. BARKER

I

IF you know something, then what you don't know won't hurt you—there is always a way of coming to know what you don't know without sacrificing the knowledge you already have. Knowledge, we might say, is of its very nature indefinitely extendible.¹ At first blush there may appear to be exceptions to this principle. Tom Grabit stole a book from the library, and Smith knows it, for he knows Tom well and he witnessed the foul deed. One thing Smith doesn't know is that Tom's mother has averred that Tom has an identical twin brother who was in the library at the time. But another thing Smith doesn't know is that Tom's mother is demented, and the twin is only a figment of her imagination.² Is it true that what Smith doesn't know won't hurt him? Could he somehow acquire knowledge that Tom's mother said what she said without suffering the consequence of having his belief that Tom stole the book disqualified as knowledge? Yes, simply by acquiring the additional knowledge as part of the whole story, including the facts about the woman's deranged mind. In getting the apparently contravening evidence as part of the whole truth of the matter, Smith could preserve his original knowledge that Tom stole the book.

Can this extendibility feature of knowledge be incorporated into a justificationist analysis of knowing? Let us hypothesize that a justified

belief in a true proposition *p* qualifies as knowledge if and only if what isn't known won't hurt, i.e., if and only if there is some way that any other true proposition besides *p* could come to be justifiably believed without destruction of the original justification for believing *p*. In other words, one who justifiably believes a conclusion has knowledge just in case there is no truth one could not somehow learn while still justifiably believing the conclusion solely on the same basis as before. When the justification for *p* is thus compatible in principle with justified belief in every other true proposition, we can say that the justification for *p* is *absolute*. Knowledge, according to this theory, is absolutely justified true belief.³

The absolute justification theory gains intuitive appeal from the following kind of consideration. If I believe that I know that Tom stole the book, then I believe that at least in the long run the evidence would support my view; any evidence that may crop up against Tom's stealing the book could, I feel sure, eventually be discredited by more evidence. According to the theory under consideration, believing that one knows entails believing that one's justification is in principle compatible with justified belief in every other true proposition. Hence, the theory succeeds in accounting for the special kind of confidence associated with my believing that I know that Tom stole the book.

¹ Risto Hilpinen, in "Knowledge and Justification," *Ajatus*, vol. 33 (1971), pp. 7-39, discussed at some length a stronger version of the extendibility of knowledge thesis, a version stating in effect that no matter how one comes to know what one doesn't know one's existing knowledge is preserved (see pp. 25 ff.). In this connection, see also: Jaakko Hintikka, *Knowledge and Belief* (Ithica, 1962), pp. 20-21; Gilbert Harman, *Thought* (Princeton, 1973), Ch. 9; and Marshall Swain, "Epistemic Defeasibility," *American Philosophical Quarterly*, vol. 11 (1974), pp. 15-25, pp. 19 ff.

² This example, which was discussed by Hilpinen in relation to the extendibility thesis, is borrowed from Keith Lehrer and Thomas Paxson, "Knowledge: Undefeated Justified True Belief," *The Journal of Philosophy*, vol. 66 (1969), pp. 225-237, see p. 228.

³ The existing theories most similar to this theory are those of Peter Klein, "A Proposed Definition of Propositional Knowledge," *The Journal of Philosophy*, vol. 68 (1971), pp. 471-482, and Risto Hilpinen, *op. cit.* Their theories claim in effect that one's justified true belief qualifies as knowledge if and only if no matter how one attains a justified belief in any other true proposition one's original justification is preserved. For criticisms of these stronger theories, see Swain, *ibid.*, and Keith Lehrer, *Knowledge* (London, 1974), pp. 22 ff. The superiority of the weaker theory is demonstrated in § 4 below.

II

In turning to troublesome cases of knowledge and of ignorance discussed in the literature, we find that the absolute justification theory has a great deal of plausibility. Many of these cases center on justifications involving false assumptions. Smith, justifiably believing that a certain colleague, Jones, owns a Ford, correctly infers that at least one of his colleagues owns a Ford, a proposition Smith thereby acquires justification for believing. It so happens that while this proposition is true (owing to the fact that another colleague, Brown, owns a Ford), the assumption that Jones owns a Ford is false. Despite having a justified belief in the true proposition that at least one of his colleagues owns a Ford, Smith obviously lacks knowledge.⁴ The absolute justification theory can easily handle this case. What Smith doesn't know is that Jones owns no Ford, and this hurts—there is no way that Smith could acquire justification for believing this true proposition without destruction of his original justification for believing that at least one of his colleagues owns a Ford. Even in light of the whole story, which includes the information that Brown owns a Ford, Smith's original justification would remain utterly devastated, despite the fact that in getting the whole story Smith would then have an independent source of knowledge that at least one colleague owns a Ford. To say that Smith's original justification would be destroyed is to say that he could no longer justifiably believe his conclusion solely on the same basis as before. Thus, acquisition of new support for his conclusion could not prevent destruction of his original justification.

The theory, however, is not so restrictive that it requires *all* assumptions to be true. Let us modify the above example and suppose that Smith is also justified in believing the true proposition that Brown owns a Ford. And let us suppose that in addition to drawing his conclusion that at least one colleague owns a Ford from his premiss about Jones, Smith independently draws this conclusion from the premiss that Brown owns a Ford.⁵ This time the falsity

of the assumption about Jones does not prevent Smith from gaining knowledge of his conclusion, and this time the absolute justification theory does not deny Smith knowledge. What Smith doesn't know won't hurt, for the adequacy of his justification for believing his conclusion would be unaffected by acquisition of a justified belief in the true proposition that Jones does not own a Ford. Despite the destruction of a portion of Smith's total justification, enough remains to justify belief in the conclusion.

Some troublesome cases involve false assumptions which are more-or-less implicit. Smith, coming across Jones' valid title to a Ford, acquires a justified belief that Jones owns a Ford, something which is true. What Smith doesn't know is that the title he saw was to a Ford which was recently destroyed by a bomb, and that Jones happened to acquire another Ford in a raffle. Obviously Smith does not have knowledge that Jones owns a Ford.⁶ It would be difficult to maintain that there was any explicit assumption of Smith's that was false. Nevertheless what Smith doesn't know hurts—he could in no way acquire a justified belief in the true proposition that the title he saw was to a car which no longer existed without suffering destruction of his original justification for believing that Jones owns a Ford.

Sometimes assumptions are at most *virtual*. Smith, let us suppose, has various items of evidence that several of his colleagues, Jones, Brown, Black, and White, each own Fords. In each case the evidence is not quite sufficient to justify a belief that the colleague in question owns a Ford. Smith also has some evidence (the fact that Fords are popular cars) that other colleagues besides these four might very well own Fords, but this evidence is not sufficient to justify a conclusion that at least one colleague owns a Ford. Nevertheless *all* of the evidence taken together is sufficient to justify this conclusion, and Smith makes the inference. Unknown to Smith, however, none of the four above-mentioned colleagues owns a Ford. It so happens that Smith's conclusion is true because another colleague, Green, does own a Ford. It is obvious that Smith

⁴ Cf. Edmund Gettier, "Is Justified True Belief Knowledge?", *Analysis*, vol. 33 (1963), pp. 121-123.

⁵ Cf. Keith Lehrer, "Knowledge, Truth and Evidence," *Analysis*, vol. 25 (1965), pp. 168-175, p. 170.

⁶ Cf. John Turk Saunders and Narayan Champawat, "Mr. Clark's Definition of 'Knowledge'," *Analysis*, vol. 25 (1969), pp. 8-9, and Brian Skyrms, "The Explication of 'X Knows that p'," *The Journal of Philosophy*, vol. 64 (1967), pp. 373-389, p. 383.

lacks knowledge.⁷ Nevertheless, it would be difficult to maintain that any *actual* assumption of Smith's was false—the false proposition that either Jones or Brown or Black or White owns a Ford constitutes at most what could be called a *virtual* assumption. But the fact remains that what Smith doesn't know hurts, for there is no way he could acquire a justified belief in the true proposition that none of these four colleagues owns a Ford without undermining his original justification for believing his conclusion.

III

In some cases of belief-acquisition a false proposition is involved not as an assumption, but as what can be called a *presupposition*. Consider the following variation on the Grabit story, which we shall dub the Second Version. Having witnessed the theft of the library book, Smith is justified in believing that Tom, whom he knows well, stole the book. What Smith doesn't know, however, is that Tom really does have an identical twin brother who happened to be in the library at the time. In this case, it seems, Smith fails to acquire knowledge that Tom stole the book.⁸ It would be implausible, however, to claim that the proposition that there was no such twin in the vicinity was an *assumption* (implicit or virtual) that Smith made, for the *absence* of belief in this proposition did not in any way affect Smith's concluding that Tom stole the book. It is more appropriate to characterize the proposition as a *presupposition* of Smith's—only the *presence* of belief in the *falsity* of the proposition would have adversely affected Smith's concluding that Tom stole the book.⁹

According to the absolute justification theory, the falsity of Smith's presupposition prevents him from attaining knowledge. There is no way that Smith could acquire justified belief in the fact that Tom's identical twin was in the vicinity without wreaking havoc on his original justifica-

tion for believing that Tom stole the book. At first it may seem that there might be a way. After all, the rest of the story is that the twin was quietly studying epistemology in another part of the library. Perhaps if Smith learned the whole story, including not only the information about the presence of the twin in the vicinity, but also the information about the twin's precise location at the relevant time, then his original justification would not be adversely affected. But this won't wash. Smith's original justification would be rendered inadequate in itself by acquisition of justified belief in the whole story, despite the fact that some of the new information, i.e., that concerning the precise location of the twin, could be used to *supplement* the original evidence in such a way as to "re-justify" the belief that Tom stole the book. The fact remains that upon learning of the presence of the twin in the vicinity, Smith could no longer justifiably believe his conclusion solely on the same basis as before. What Smith doesn't know hurts, and hence the absolute justification theory rightly entails that Smith lacks knowledge.

IV

To see that the falsity of a presupposition does not always prevent one from attaining knowledge, let us return to the First Version of the Grabit story. Smith, who witnessed the crime, is justified in believing that Tom, whom he knows well, stole the book. Tom is indeed the culprit, but unknown to Smith, Tom's mother has averred that Tom has an identical twin brother who was in the library at the time. No doubt if Smith, who had no thoughts at all about such a twin, had believed that Tom's mother said this, he would have at least hesitated to conclude that Tom stole the book. Hence Smith can be said to have *presupposed* that there was no such claim made. But this time the falsity of his presupposition does not prevent his attaining

⁷ This type of example is suggested by considerations advanced by Lehrer in *Knowledge*, *op. cit.*, pp. 20, 218.

⁸ Cf. Klein, *op. cit.*, p. 474.

⁹ The question of the relevance of such presuppositions to knowledge is closely associated with the question of the relevance of unfavorable evidence which the subject does not possess; the latter is discussed by Ernest Sosa in "Analysis of 'Knowledge that p'," *Analysis*, vol. 25 (1964), pp. 1-8, p. 5, and in "Two Conceptions of Knowledge," *The Journal of Philosophy*, vol. 57 (1970), pp. 59-66, p. 62; and by Gilbert Harman in "Knowledge, Inference, and Explanation," *American Philosophical Quarterly*, vol. 5 (1968), pp. 164-173, pp. 172-173, and in *Thought*, *op. cit.*, Ch. 9. Related topics are discussed by Fred Dretske in "Epistemic Operators," *The Journal of Philosophy*, vol. 57 (1970), pp. 1007-1023; by Alvin Goldman in "Discrimination and Perceptual Knowledge," presented at the Univ. of Missouri at St. Louis, 1974; and by G. C. Stine in "Skepticism, Relevant Alternatives, and Deductive Closure," presented at the 1974 meeting of the American Philosophical Association, Eastern Division.

knowledge, for the twin is only a fiction of the woman's deranged mind. And the absolute justification theory does not deny Smith knowledge. If Smith should acquire justified belief in the fact that Tom's mother said what she said along with the information about her delusions, his original justification would be preserved intact. The facts about the woman's mental condition render her claim incapable of damaging Smith's justification for believing that Tom stole the book. Upon learning the whole story, Smith could justifiably continue to believe his conclusion solely on the same basis as before. Thus, the absolute justification theory is not so restrictive that it requires all presuppositions to be true.

At this point we should consider the following possible objection. Suppose that it is true that what mothers say in situations like that described above is generally reliable. Now this fact, coupled with the fact that Tom's mother said what she said, has the consequence that Smith does *not* know that Tom stole the book—it is only a felicitous coincidence that Smith's belief is true, and hence he lacks knowledge.¹⁰ But the absolute justification theory fails to deny Smith knowledge, for there is a way, via learning the above-mentioned facts along with the facts about this mother's demented state, that Smith could extend his fund of justified true beliefs without undermining his original justification. Hence, the theory is too weak, as it classifies a case of ignorance as a case of knowledge.

But is this really a case of ignorance?¹¹ Granted, there are, unknown to Smith, some facts which if learned would constitute strong evidence against Tom's being the culprit. This evidence, however, would be misleading evidence; and it could be completely nullified by additional evidence concerning the woman's mental condition. The mere possibility of misleading evidence seems insufficient to prevent Smith's belief from qualifying as knowledge, since the misleading character of the evidence could be discovered by further investigation. The above objection presupposes that one cannot know unless there is no possibility of misleading evidence, i.e., unless there are no true

propositions which would if known constitute strong unfavorable evidence against the true proposition one believes. Such a requirement seems too stringent to accord with the ordinary concept of knowledge. For instance, when one believes that one knows, one does not necessarily believe that no misleading evidence could crop up; one need hold only that *if* apparently contravening evidence does come to light, then this evidence could eventually be nullified by additional evidence. If one is convinced that one knows, then one does not fear the whole truth of the matter, though one might very well fear partial truths. Thus the claim that Smith would fail to attain knowledge in the situation described above rests on an erroneous view concerning the requirements of knowing. The absolute justification theory is correct in according knowledge to Smith.

V

We have seen that with regard to the First Version of the Grabit story, in which it is reasonably clear that Smith attained knowledge, the absolute justification theory is in agreement with preanalytic judgment. And with regard to the Second Version the theory likewise accords with intuition in denying Smith knowledge. By modifying the First Version we can see that as it becomes less clear that Smith attains knowledge, it becomes correspondingly less clear that Smith's justification is absolute. For example, instead of supposing that Tom's mother is demented, let us suppose that she is quite sane, and fabricated the twin story because she wanted to protect Tom, whom she feared might have been the culprit.¹² It is no longer very clear that if Smith should learn of her claim along with the facts about its real source, his original justification would remain unaffected. But neither is it very clear from the intuitive point of view that Smith would have knowledge that Tom stole the book. If we introduce the further supposition that Smith did not himself witness the crime, but learned of it from a newspaper account, we are more strongly inclined to deny that Smith attained knowledge.¹³ At the same time, however,

¹⁰ Cf. Klein, *op. cit.*, pp. 479–480, Footnote 9; and Hilpinen, *op. cit.*, pp. 32 ff.

¹¹ Lehrer, discussing a similar case, argues that knowledge is attained by the subject (see *Knowledge, op. cit.*, pp. 221–222). In this connection see also Swain's discussion of the Rock Throwing Example, *op. cit.*, p. 20.

¹² Cf. e.g. Sosa, "Two Conceptions of Knowledge," *op. cit.*, p. 62, and Harman, *Thought, op. cit.*, p. 142.

¹³ Cf. e.g. Harman, *ibid.*, pp. 143–144.

we are more strongly inclined to say that in this case Smith's justification is not absolute—if he should acquire a justified belief that Tom's mother said what she said, he would perhaps suffer destruction of his original justification, even if he learned of this claim in conjunction with the facts about its real source. Justification stemming from communication, it seems, is more sensitive to contravening evidence than that attained through perception.

In a similar fashion we can modify the Second Version of the Grabit story to show that as it becomes less clear that Smith failed to attain knowledge, it likewise becomes less clear that Smith's justification would be adversely affected by an expanded fund of justified true beliefs. Instead of supposing that the twin was in the library at the relevant time, let us suppose that he wasn't there because he was involved in an automobile accident on his way to the library. Did Smith fail to attain knowledge that Tom stole the book? Could Smith have learned of the existence and intended destination of the twin without adverse effect on his original justification, provided that he also learned about the accident? Clear-cut answers do not seem available to either question. Let us go on to suppose that the accident occurred a thousand miles away three days prior to the theft, and that the twin was killed in the accident.¹⁴ Again the absolute justification theory accords with intuition—both preanalytic judgment and the theory seem less disposed in this case to condemn Smith to ignorance.

These considerations indicate that the sphere of vagueness of the ordinary term "know" may turn out to coincide with the sphere of vagueness of the term as defined by the absolute justification theory. Such a congruence would constitute additional evidence in favor of the theory.

VI

If knowledge is absolutely justified true belief, then is the skeptic right about the extent of our knowledge—is he justified in believing that we possess little or no knowledge? Typically, the skeptic focuses on certain general presuppositions of our conclusions. For instance, the skeptic may note that in coming to almost any con-

clusion we presuppose that there is not some evil genius manipulating our beliefs, for if we believed there were such a being operating behind the scenes we would hesitate to draw the conclusion. The skeptic inquires as to whether we are justified in believing that this presupposition is true; failing to find sufficient justification to satisfy himself, he concludes that he is justified in believing that we fail to attain knowledge. Now according to the absolute justification theory, we do not attain knowledge unless the above-mentioned presupposition is true. But the theory of knowing the skeptic implicitly appeals to is much more restrictive—according to his theory, even if the presupposition is true, we fail to attain knowledge unless we are *justified* in believing the presupposition.

If we accept the skeptic's theory of knowing, then skepticism may appear to be a reasonable thesis about the extent of our knowledge. Our investigation of the absolute justification theory, however, points to the conclusion that the skeptic's theory is too stringent to constitute an accurate explication of the ordinary concept of knowledge. According to this stringent theory, our presuppositions must be justified as well as true in order for us to have knowledge. But, it seems, the ordinary concept of knowledge requires only that the presuppositions be true. (Indeed, as we have seen in the First Version of the Grabit story, sometimes the downright falsity of a presupposition is incapable of preventing attainment of knowledge.) Whether or not our presuppositions are justifiably believed is a question relevant not to our having knowledge, but to our having knowledge that we have knowledge. The skeptic has no right to assume that we don't know unless we know that we know.

If the skeptic can show that we are not justified in believing the general presuppositions of our conclusions, then he will have shown that we are not justified in believing that we know. But according to the absolute justification theory, he will not have thereby shown that he is justified in believing that we do not know. Although skepticism may turn out to be the truth about knowledge, the absolute justification theory seems to provide no aid or comfort to the skeptic. On the other hand, the theory is not by any means a boon to the dogmatist, as it does not facilitate the dogmatist's task of showing that

¹⁴ Cf. e.g. Harmon, *ibid.*, p. 143, and Goldman's discussion of the Barn Example in *op. cit.*

he is justified in believing that we have a good deal of knowledge. Indeed, the theory seems just right, allowing the skeptic and the dogmatist at least to agree on what it is they disagree about.¹⁵

VII

Knowledge is in principle indefinitely extendible. The absolute justification theory of knowing, which is based on this principle, turns out to exhibit considerable promise as a satisfactory analysis of the ordinary concept of propositional knowledge. Not only can the

theory discriminate between damaging and harmless instances of false assumptions, it can do the same with respect to instances of false presuppositions. The applicability of the concept of knowledge defined by the theory seems clear where the applicability of the ordinary concept of knowledge is clear; and where the one concept is vague, the other is also. Finally, the apparent doctrinal neutrality and pragmatic adequacy of the theory with respect to the perennial quest for the bounds of human knowledge lend strong support to the view that knowledge is absolutely justified true belief.¹⁶

Southern Illinois University at Edwardsville

Received June 24, 1975

¹⁵ Cf. Klein, *op. cit.*, pp. 481-482.

¹⁶ I am indebted to Thomas Paxson and Kenneth Collier for helpful comments on an earlier version of this paper.

VII. IRONIC OR NOT?

CLYDE LEE MILLER

IN this paper I want to consider the word "ironic" as it occurs in ordinary language. In conversational English "ironic" appears most often in such comments and questions as "It was ironic that X." "She was (is) being ironic (in saying or doing X)." "Ironic, wasn't (isn't) it?" In such utterances, the X said to be ironic is part of a sentence about some situation, occurrence or locution. Such employment of "ironic" might raise the following questions. What entitles us to characterize such cases as ironic? Does our language reflect a viewpoint we adopt or is something else going on, independently of how we look at things? When ought we challenge someone's saying "X is ironic?"

Perhaps we are most familiar with examples of verbal irony—where a person speaks ironically. But I want to deal first with the non-verbal occurrences we correctly term ironic. Speaking ironically may be better understood over against non-verbal irony—where what is ironic stands apart from intention and utterance of any speaker.¹

I

I begin with a list of examples of ironic situations. Each example could be correctly introduced with the words "it is ironic that..." where the example would complete the sentence. The list is intended to display what we note about the world when we say that something is ironic.

1. The wealthiest woman in the country club won the state lottery last week.
2. The more time I put in the less work I get done.
3. Most blind dates she had were real "dogs" so she did not want another one. She got her roommate to go instead and he was this beautiful guy.
4. He always shouts at his children to get them to be quiet.

In each of these examples some more or less obvious reversal could startle anyone who noticed what happened. The outcome or course of events we would take as typical or sensible in each case comes out the other way around. Each example affords some pointed contrast between the participant's or observer's sense of what is fitting or proper in such circumstances and what actually transpired. This contrast could be spelled out to state what we generally consider appropriate in the case at hand and then to state what in fact did take place. The first statement would run counter to the second.

But is this precise enough to distinguish what is ironic from what is just extraordinary? I may not expect blind dates to be remarkably attractive or unattractive. When one is either especially pleasing or displeasing, my "realistic" expectations may be countered, even drastically. But I can hardly term this good or bad luck *ironic*—and I should challenge anyone who called it ironic as missing the point. We may assess a situation as odd, remarkable, unusual, or atypical, but for it to be ironic more is involved.

Again, winning the state lottery is not something any one person counts on happening to her. Just this statistical improbability enables us to see any winner as fortunate or at least unexpectedly singled out. Yet if someone called this eventuality "ironic" for a typical purchaser of lottery tickets, I would be at a loss to know what she saw in the situation that I had missed.

The contrariety in an ironic situation or eventuality does counter the habitual expectations and received understandings of those involved in or observing the occurrence. (Thus we expect more time expended to yield greater output, other things being equal.) Such expectations and beliefs are based on what agents or observers believe they are doing or is being done, or are trying to accomplish, or think fitting or not, or can reasonably expect in such circumstances. But contrariety or mere reversal is not enough

¹ Robert Fogelin of Yale University first recommended this strategy to me. For a fuller treatment of the same distinction in aesthetic and literary contexts, see D. C. Muecke, *The Compass of Irony* (London, 1969), pp. 42-52.

for irony. Outcomes that are unfortunate, disappointing, or even tragic may manifest such reversals and not be ironic, just as successful or happy results may be contrary to expectations and in no way be ironic.

What is distinctive of ironic situations is that the inappropriate outcomes, while they run contrary to the hopes and efforts of those involved, display a particular propriety as well. What eventuates may counter conventional or normal expectations, but it also functions as the "just deserts," whether good or bad, of those attempting, awaiting or just hoping for and believing in the opposite. Normal outcomes have been thwarted, only to have "poetic justice" (which may prosaically reward as well as punish) done. What turns out contrary to conventional wisdom and habitual belief is quite fitting when measured in some wider context of beliefs or facts relevant to those acting in or observing the situation.

When the woman who wins the lottery is rich, not only does she not expect this good fortune any more than others do, but *she* may not even find it economically desirable and so her windfall may be taken as doubly misplaced. Our spare economic conditions remain unaltered while she does not need such winnings. Little wonder *she* was the winner. In the earlier example, the refusal to extend past bad luck with blind dates rewards one's roommate with good luck. Here expectations are countered because such precautions against bad fortune are too secure; they also prevent any good outcome for the person taking them. Such additional circumstances give ironic occurrences their appropriate impropriety—outcomes where what the principals get illuminates what they deserve no matter what they want or attempt.²

The remarks we employ to reject another's appraisal of a situation as ironic underline what is distinctive about such ironies. We would refuse to call some turn of events ironic when no reversal or contrariety can be found in it. To such misuses we might say: "That's not ironic, he planned on it." "Why, it's just what she feels she deserves." "What else would you expect?"

Such refusals are appropriate because an eventuality that does not counter efforts and hopes cannot qualify as ironic to begin with. (At best a rejection of "ironic" in such cases invites further explanation from one who so appraises them.)

But we may also reject the use of "ironic" even where some reversal of fortune or expectation does occur. What justifies such rejection is the fact that some reversals, though odd or misplaced or even lucky, have little or no connection with the efforts, beliefs, attitudes or dispositions of the participants in or observers of the affair in question. Our responses to such misuses of "ironic" change accordingly: "It can't be ironic, there is no way this turn of events was deserved." "Ironic? Hardly; it is simply bad luck." "Of course it is disappointing (encouraging)—but why do you call it ironic?"

All such refusals remind us that whether an event or set of circumstances is ironic or not is a function of more than our calling it such or wanting it to be so. Because of our experience with literature and its twentieth-century critics, we are tempted to believe that for any situation a perspective can be found from which everything (or nothing) is ironic. But this confounds discovery and invention, overlooking the fact that literary situations are deliberately constructed. (Of course one can treat life as if it were literature, but the "as if" undercuts the validity of the move—and underlines its sadness.) In non-literary contexts no one quite plays the role of author, so the possibilities for irony are significantly lessened. In ordinary experience where people speak and act we can observe an unexpected turn of events and be mistaken in terming it ironic.³

The word "ironic" belongs among those adjectives we employ in judging and appraising how and whether the world of ongoing human endeavor fits our usual notions of the way things should be. We assess circumstances and situations in which we participate as agents and observers, judging them fitting and proper, right and just—or else incongruous and out of joint. The ways in which events go awry lead us to employ such

² Eddy Zemach of S.U.N.Y. at Stony Brook got me to revise several points here by proposing that an ironic contrariety has some "particularly illuminating serendipity" to it. (I am also grateful to Karsten Harries of Yale University and to David Kolb of the University of Chicago for comments on an earlier version of this paper.)

³ Regarding irony in literature, see Muecke, *Compass*, *op. cit.*, and his shorter *Irony* (London, 1970) in "The Critical Idiom" series; also Wayne C. Booth, *A Rhetoric of Irony* (Chicago, 1974). All three books contain extensive bibliographies.

words as "comic," "tragic," "paradoxical," "anomalous," "extraordinary," "ironic." "Ironic" thus takes its place among related words we might term assessment adjectives, or at least among that group of adjectives used to appraise untoward and unexpected occurrences. "Ironic" is employed to match that special lack of fit in situations where we recognize that we (or others) are getting the opposite of what we anticipate as or believe to be appropriate, yet what we truly deserve when we recall or note further aspects of the situation.

Situations and circumstances we should call ironic thus display a characteristic tripolar structure. At one pole stands what those engaged in the situation believe is being done, what they are attempting, hoping for, or see as fitting. The second and opposed pole may be construed as what actually occurs or is the case. Those involved at the time often are aware of this opposition, though they need not be. It is crucial that what in fact happens takes place instead of what was anticipated or believed and counters it. Only if this opposition is present can a situation qualify as ironic; this bipolar opposition may be taken as the necessary condition for a situation to qualify as ironic.

But more is required for irony. The countering outcome or happening must display withal an appropriateness to the total situation and those involved, a characteristic fit in spite of its contrariety—this is the third pole required for events to be ironic. The implicit evaluation that what happened is fitting nonetheless matches this third component in the structure of irony. This threefold structure goes beyond mere polar opposition, for it is the suitability in the countering facts or events which counts as the sufficient condition for a situation or occurrence to be truly ironic.

II

A. I turn next to *being ironic* in the sense of speaking ironically, to the occasions where a person is ironic in what he or she says. Again contrariety is present, at least frustration of the hearers' habitual expectations, but now within the speaking itself. Ironic locutions display a special twist absent from non-ironic utterances.

The latter lead us to expect that speakers say what they mean and mostly mean what they say—even in metaphorical or other non-literal speaking. But ironic speaking seems an instance where we must not mean what we say. As the dictionary has it, in irony "the intended meaning is the opposite of that expressed by the words used."⁴

Suppose, for example, that I have just been seated for dinner at the home of old friends. With some embarrassment my host, an earnest chef, serves a generous sirloin steak—quite overdone. As I saw through the charred beef, I smile and say gently, "Another triumph, my friend."

Given the situation and the condition of the steak, my remark may be inappropriate on several counts. The face value of the comment belongs to a different occasion, for the explicit sense of my words should fit a moment of culinary success. Suppose another guest tried to explain: "Oh, he's just being ironic." "What makes you think so?" "I doubt if he means quite what he said." Perhaps my remark could count as ironic on the vague grounds that my words do not or cannot or should not mean what they say, or rather, that what they do mean may not match my intent in uttering them or fit the situation. Can these hints be given a more precise explanation?

It is possible to expand the initial remark so that it reads less elliptically, thus: "This (situation, dinner, beefsteak) is another triumph." If this utterance is analyzed in terms of its propositional content (its sense and reference), it does refer to or pick out the dinner's meat course and its impact on the affair. But if the usual sense or conventional meaning of "triumph" is recalled, there is nothing to which this word could apply as a predicate.⁵ I am speaking of a state of affairs such that a different remark using some predicate opposite in sense is called for, for example, "Too bad." "What a shame." Generally it is through our understanding of the sense of a locution that we are able to apply it in a given case. Speaking ironically, however, sets up this incongruity: we use an expression whose straightforward or primary sense—here, that of its predicate—does not apply, and we use it precisely to refer to and assess the situation at hand.

"Another triumph" did refer, in fact, to the

⁴ *Oxford English Dictionary*, s.v. "Irony."

⁵ My use of "propositional content" attempts to follow J. R. Searle, *Speech Acts* (Cambridge, 1969).

blackened beef and assess its effect on the occasion. But if one were to say anything at all, "What a disaster" would be closer to the utterance the situation *prima facie* calls for. Because the sense of "disaster" runs counter to that of "triumph," this opposition or contrariety suggests that an ironic utterance may be another case of an ironic situation. In the present example, triumph is fairly conspicuous for its absence at the dinner; indeed, it may be failure's too evident presence which enables my use of "triumph" to speak of this situation at all.⁶

The point of the contrariety in ironic speaking is more than a matter of the speaker's intentions and purposes, however wide-ranging these may be. Or better, what is intentional in ironic locutions, that is, what the speaker is trying to do in such speaking, turns out to be a matter of intension or sense as well—a matter of the conventional meaning or the propositional content in the utterance. One crucial element in an ironic locution consists in the opposition of sense between the expression(s) typical or expected in such a case and the expression actually uttered to speak of it. But does this opposition differentiate ironic speaking from any comment out of place in the circumstances, such as a verbal slip or puzzling misinterpretation?

My response to the charred beeksteak, for instance, could have been incongruous or out of place without being ironic. Had I commented on the size of the portions or the bouquet of the wine, I might have eased my host's embarrassment (or just added to it). Even if just my speaking up were taken as ironic, there was nothing ironic in *what* I was saying.

The incongruity of ironic locutions lies in *what* is said—the predicate expressions employed do not apply to the situation at hand but to some opposite state of affairs. Yet this discrepancy is not out of place as if the speaker were simply mistaken or even lying; an ironic locution is also fitting and apposite since hearers can understand what is being spoken about and what is being said about it—the opposite of what the

words conventionally mean. As hearers we recognize what the speaker is getting at, now in spite of *and* in virtue of the expression uttered. What enables us to grasp the point and fit of ironic speaking is that such locutions render some (usually negative) evaluation or assessment in a situation where both speaker and hearers expect such attention to be paid. Both can be delighted and enlightened at finding their expectations fulfilled in terms that apparently make the opposite assessment. What makes this opposite sense apposite is that its expression provokes another or closer look at what is spoken of, the situation, and those involved in order to measure the locution's proffered assessment and to be able to reject its face value.⁷

In ironic locutions, then, we find a tripolar structure not unlike that earlier remarked in ironic eventualities. As the latter manifest an appropriate impropriety once the countering turn of events is taken in a wider context, so ironic speech displays its own apposite opposition. But since the opposition is a matter of the sense of the locution actually used, its appositeness consists in the way such speaking reveals for speaker and hearers a shared (and often unfavorable) evaluation of the matter spoken about. At least in the simplest ironic speech what is agreed on generally can be fittingly appraised in just the wrong terms. The shared implicit assessments of speaker and hearers make up the third pole beyond the opposition within the utterance's propositional content.

B. Ironic utterances may also have illocutionary and perlocutionary force.⁸ In speaking ironically I may also carry out other purposes—from smoothing over an uncomfortable situation to expressing a wry appreciation of what might have been. My host at the dinner may have complained that my "triumph" comment was sarcastic; had I rejoined that I was being ironic, I may not have been excused. But this is not just because being ironic and being sarcastic have similar connotations and often occur together in everyday usage. Sarcasm is a different issue, one

⁶ Ted Cohen extends this point to whole speech acts: "Just as with some (but not all) figurative sentences, the fact that taken literally they are transparently false is the key to recognizing that they are figurative, so it is with whole speech acts: the fact that it is transparent to all that the associated perlocution is out of the question is the key to recognizing that the conventional force is not to be actualized." "Illocutions and Perlocutions," *Foundations of Language*, vol. 9 (1973), p. 503.

⁷ H. P. Grice underlines the evaluative aspects of ironic speech in his unpublished William James lectures, "Logic and Conversation," 1966. See lecture III.

⁸ J. L. Austin, *How to Do Things with Words* (New York, 1965); J. Searle, *Speech Acts*, *op. cit.*; T. Cohen, "Illocutions and Perlocutions," *op. cit.*

of the force of a locution, of what I intend to do in saying what I say.

To distinguish irony from sarcasm, it may help to state a generalization, then use examples to explain it. An ironic utterance can be employed sarcastically, but ironic speaking as such is a matter of the lack of fit in the sense of the words used, whatever else one may be doing in and through such speaking. To be sarcastic one need not be ironic. The former is a matter of an utterance's illocutionary force, the latter of its propositional content.

Return once more to the *beefsteak*. "Another triumph" sounds like a fairly nasty remark; it could hurt feelings even more effectively if spoken without the cutting tone of voice we usually hear as sarcastic. What makes it sarcastic, in part, is my intent to derogate or offend my host. Even if "Another triumph" were a thoughtless remark or a joke that misfired, it could end up wounding by suggesting the scornful superiority one often assumes in snide comments. Yet the fact remains that I was not forced to use an *ironic* utterance if I did want to be cutting and derogatory.

Instead I could have said cuttingly, "Done to a burn, eh?" In saying this not only was my intention sarcastic, but my words had a sense to match the precise state of the *beefsteak*. I now deserve the retort: "You meant exactly what you said and you were trying to be nasty." Here the most effective sarcasm was to underline what's what by describing it exactly. That we can also be sardonic or sarcastic or snide by saying the opposite of what's what (i.e., by speaking ironically) is incidental to identifying the presence of sarcasm.

What the distinction of an utterance's illocutionary force from its propositional content or meaning does is show how ironic speech can carry off other intents and purposes besides foiling the listener's mind-set for straightforward language usage. I propose the following list of examples to indicate the variety of possible forces with which irony can be employed.

1. sarcasm: I pick up every toy in the room, turn to the child, and say: "Thanks a lot."

2. playfulness: I come in with freering rain pouring off me and say cheerfully: "What a lovely day."
3. teasing resignation: I get a fourth distinct sum on my fourth attempt to add a column of figures. My friend chuckles: "I think you're getting it."
4. unintended, self-revealing: The Pope signs his letters: "The servant of the servants of God."

It is easy to see how ironic and sarcastic talk get identified in common parlance and conventional wisdom since each example could be employed sarcastically to offend without much alteration. We might view ironic speaking as a special way of being sarcastic, at least on the conversational evidence of adult experience. We seem much harder put to find examples of remarks which are at once ironic and constructive or favorable, rather than ironic and derogatory. Without the implied negative criticism, ironic remarks turn out to be a kind of joke. But to keep clear the meaning and use of "ironic" as it applies to speaking we do better to take sarcastic irony as one sort of ironic locution. When the range of illocutionary forces is added to the contrariety central to an ironic locution, the possible combinations and complications which attend ironic speaking are many indeed.

Yet we seldom use the term "ironic" in an ironic way. "Ironic" is most often employed in non-ironic assessments and appraisal of other utterances or situations. Whether what is evaluated is in fact ironic or not, the locutions "It is ironic that X" and "She was being ironic (in saying X)" are direct and straightforward. Now we can and do say "Isn't X ironic?" ironically, meaning that X is not ironic at all. But this possibility witnesses to regular usage. Not only would talk get incredibly tangled if "ironic" (or almost any other word) were used frequently in ironic fashion, but the point and wit of ironic utterances would be lost as well. Ironic locutions themselves fall flat when they are taken straightforwardly—thus children foil their parents' verbal cleverness. Such is not the case with straightforward talk; what is already on the ground has no place to fall.

BOOKS RECEIVED

- Abel, Reuben, *Man is the Measure* (New York: The Free Press, 1976), 296 pp.
- Bastable, Patrick K., *Logic: Depth Grammar of Rationality* (Atlantic Highlands, N.J.: Humanities Press, 1976), 429 pp. \$32.50.
- Beauchamp, Tom L. (ed.), *Ethics and Public Policy* (Englewood Cliffs: Prentice-Hall, Inc., 1975), 442 pp.
- Bontempo, Charles J. and S. Jack Odell (eds.), *The Owl of Minerva* (New York: McGraw-Hill, 1975), 264 pp. \$3.95.
- Capaldi, Nicholas, *David Hume* (Boston: G. K. Hall & Co., 1975), 241 pp. \$7.50.
- Culbertson, James T., *Sensations, Memories and the Flow of Time* (Santa Margarita, Calif.: The Cromwell Press, 1976), 189 pp. \$8.00.
- De Luca, Anthony J., *Freud and Future Religious Experience* (New York: Philosophical Library, 1976), 261 pp. \$12.50.
- Hintikka, Jaakko (ed.), *Rudolf Carnap, Logical Empiricist* (Dordrecht: D. Reidel Publishing Co., 1975), 400 pp.
- Lehrer, Keith (ed.), *Analysis and Metaphysics* (Dordrecht: D. Reidel Publishing Co., 1975), 317 pp.
- Lenk, Hans, *Normenlogik* (München: Verlag Dokumentation, 1974), 255 pp.
- Mach, Ernst, *Knowledge and Error* (Dordrecht: D. Reidel Publishing Co., 1976), 393 pp.
- Mittelstrass, Jürgen, *Methodologische Probleme* (Frankfurt am Main: Suhrkamp Verlag, 1975), 264 pp.
- Porter, Burton F., *Personal Philosophy* (New York: Harcourt Brace Jovanovich, 1976), 523 pp. \$6.95.
- Poster, Mark, *Existential Marxism in Postwar France* (Princeton: Princeton University Press, 1976), 415 pp. \$17.50.
- Rescher, Nicholas, *Unselfishness: The Role of the Vicarious Affects in Moral Philosophy and Social Theory* (Pittsburgh: University of Pittsburgh Press, 1975), 128 pp. \$9.95.
- Timpanaro, Sebastiano, *On Materialism* (Atlantic Highlands, N.J.: Humanities Press, 1976), 260 pp. \$15.50.

CORRIGENDA

Volume 12 (1975)

Carl Wellman: *Ethical Disagreement and Objective Truth* (pp. 211-222).

p. 218, col. 2, line 10: read falid for valid
line 12: read leasoning for reasoning
line 13: read falid for valid
line 13: read pruth for truth
line 15: read leasoning for reasoning
line 17: read lational for relational
line 19: read pruth for truth
line 23: read pruth for truth
line 26: read pruth for truth
line 28: read leasoning for reasoning
line 31: read pruth for truth
line 32: read leasoning for reasoning
line 34: read pruth for truth
line 40: read pruth for truth

p. 219, col. 1, line 15: read pruth for truth
line 18: read leasoning for reasoning
line 33: read leason for reason
line 42: read leasoning for reasoning
line 42: read pruth for truth
line 52: read leason for reason

col. 2, line 15: read falidity for validity
line 16: read falidity for validity
line 20: read pruth for truth
line 31: read pruth for truth
line 38: read leasoning for reasoning
line 45: read leasoning for reasoning
line 50: read leasoning for reasoning
line 50: read pruth for truth

p. 220, col. 1, line 1: read leasoning for reasoning
line 6: read leason for reason
line 11: read lational for relational
line 13: read lational for relational
line 14: read leasoning for reasoning
line 15: read lational for relational
line 16: read leasoning for reasoning
line 22: read leasoning for reasoning
line 23: read leason for reason
line 29: read leason for reason
line 31: read leasoning for reasoning
line 37: read leason for reason

CORRIGENDA

Volume 13 (1976)

John Leslie: *The Value of Time* (pp. 109-122).

p. 112, col. 2, line 25: should read "absentness" for "absentees"

p. 116, col. 2, line 2: insert "than" between rather and of

Doris Olin: *Newcomb's Problem: Further Investigations* (pp. 129-134).

p. 129, col. 1, line 15: delete comma; insert comma after "by"

col. 2, line 4 from bottom: delete comma

last line: first " $>$ " should read " $\geq$ "

p. 131, col. 1, line 24: " g_1 " should read " b_1 "

line 2 from bottom: delete comma

p. 132, col. 1, line 8: "(2)" should read "(3)"

line 11: "(6)" should read "(5)"

p. 133, col. 2, line 12 from bottom: replace "in fact" by "free"

David Lewis: *The Paradoxes of Time Travel* (pp. 145-152).

p. 152, col. 1, line 2 from bottom: delete "not"

DIALOGUE

Canadian Philosophical Review
— *Revue canadienne de philosophie*

Vol. XV — No. 3

Septembre / September 1976

ARTICLES

- | | |
|--|-----------------------|
| Entre la sociologie compréhensive et la sociologie behavioriste | J. N. KAUFMANN |
| Usage logique, psychologie et idéologique de la notion de gratuité | RENÉ PELLERIN |
| The Humean Female | STEVEN BURNS |
| The Consistency of Hume's Position Concerning Women | LOUISE MARCIL LAGOSTE |
| Volition Under Hypnosis | J. T. STEVENSON |
| La structure tétrahexaédrique du système complet des propositions catégoriques | PIERRE SAURIOL |

NOTE — DISCUSSION

- | | |
|---|-------------------|
| Heidegger and Plato's Notion of 'Truth' | JOHN PHILIPPOUSIS |
|---|-------------------|

COMPTES RENDUS — BOOK REVIEWS

AUSLEGUNG

A Graduate Journal of
PHILOSOPHY

Editors:

James W. Austin
Joe D. VanZandt

Board of Advisory Editors: Richard Cole, Richard T. DeGeorge,
Anthony C. Genova, Rex Martin,
Nancy D. Simco

AUSLEGUNG is a new journal published three times a year at the University of Kansas. The journal is intended as a forum for the expression of any and all philosophical perspectives. The editors are primarily interested in publishing the work of new Ph.D.'s and students pursuing the Ph.D. degree, but all papers of philosophical interest will be considered. It is hoped that AUSLEGUNG will serve as an individuating force for those entering the crowded job market in the philosophy profession today.

Annual Subscription Rates: \$6.00/Individuals, \$4.00/students, \$10.00/institutions.

Manuscripts, inquiries and subscriptions should be addressed to:

AUSLEGUNG
Department of Philosophy
University of Kansas
Lawrence, Kansas 66045



the philosophical forum A QUARTERLY

Volume VI, No. 1
Special Issue
Marx W. Wartofsky, Editor

PHILOSOPHY & EDUCATION

Fritz K. Ringer Class, Status and Education

Carol C. Gould From the Dialectic of Questions to Social Critique:
Proposals for a Concrete Phenomenology of Education

Michael Martin Equal Education, Native Intelligence and Justice

Walter Feinberg Behavioral Theory and Educational Policy

Norman Daniels IQ, Intelligence and Educability

Courtney Cazden Language Education: Learning *that*, Learning *how*,
and Learning *to*

James F. Ross On the Concepts of Reading

Mail to: Department of Philosophy, Box 2,
Boston University, 232 Bay State Road, Boston,
Massachusetts 02215

INQUIRY

Editor: *Alastair Hannay*
Founding and Consulting Editor: *Arne Naess*

Volume 19 (1976), No. 1 (Spring)

Dualism and Bodily Movements
What is Wrong with Incest?
Spiral Dependence between Theories and Taxonomy
Reductionism and Structural Anthropology
The Training of Socrates
Marx's Analysis of Commodity Exchange: A Reply to Carver
The Other Side of Ecology in Ancient Greece: Comments on Hughes
E. Gellner: *Legitimation of Belief*

Thomas W. Bestor
Jerome Neu
Berent Enç
Ivan Strenski
Philip B. Wright
Ulrich Steinworth
John Rodman
Peter Mew

Volume 19 (1976), No. 2 (Summer)

The Foundations of Cognitive Activity
Justice and the Family
Rawls and Gandhi on Civil Disobedience
Toward a Theory of Moral Debt. Pt. 2. The Factual Grounds.
A: The Good and Human Freedom
The Durability of Impossible Objects: A Reply to Lambert
On 'The Durability of Impossible Objects'
Charles Taylor: *Hegel*
Steven Lukes: *Power: A Radical View*

Knut Erik Tranøy
Francis Schrag
Vinit Haksar
Morris B. Storer
Richard Routley
Karel Lambert
G. H. R. Parkinson
Akeel Bilgrami

Subscription: £6.25, \$15.00 (institutions): £5.00, \$12.00 (individuals)

Published quarterly by UNIVERSITETSFORLAGET, P.O. Box 7508,
Skillebekk, Oslo 2, Norway

Western Ontario Series

A Series of Books on Philosophy of Science, Methodology, and Epistemology

Published in connection with the University of Western Ontario
Philosophy of Science Programme

1. Science, Decision and Value

Edited by J. J. Leach, R. E. Butts and G. Pearce
1973, vii + 213 pp.
Cloth Dfl. 58.50 / US \$19.00
Paper Dfl. 27.50 / US \$ 9.00

2. Contemporary Research in the Foundations and Philosophy of Quantum Theory

Edited by C. A. Hooker
1973, xx + 385 pp.
Cloth Dfl. 105.- / US \$37.50
Paper Dfl. 58.- / US \$19.50

3. The Interpretation of Quantum Mechanics

by J. Bub
1974, ix + 155 pp.
Cloth Dfl. 50.- / US \$19.50
Paper Dfl. 30.- / US \$11.00

4. Contemporary Research in Philosophical Logic and Linguistic Semantics

Proceedings of a Conference held at the
University of Western Ontario, London, Canada
Edited by D. J. Hockney, W. Harper, and B. Freed
1975, viii + 332 pp.
Cloth Dfl. 100.- / US \$39.00
Paper Dfl. 60.- / US \$24.00

5. The Logico-Algebraic Approach to Quantum Mechanics

Volume I: Historical Evolution
Edited by C. A. Hooker
1975, xv + 607 pp.
Cloth approx. Dfl. 175.- / US \$68.00
Paper approx. Dfl. 85.- / US \$34.00

6. Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science

Proceedings of an International Colloquium,
The University of Western Ontario, May 1973
Edited by W. L. Harper and C. A. Hooker

Complete set of 3 volumes:

1976, xxxvi + 1004 pp.
Cloth Dfl. 280.- / US \$105.00
Paper Dfl. 160.- / US \$ 60.00

Volume 1: Foundations and Philosophy of Epistemic Applications of Probability Theory

1976, xii + 308 pp.
Cloth Dfl. 105.- / US \$39.50
Paper Dfl. 65.- / US \$21.50

Volume 2: Foundations and Philosophy of Statistical Inference

1976, xii + 455 pp.
Cloth Dfl. 130.- / US \$49.00
Paper Dfl. 70.- / US \$27.00

Volume 3: Foundations and Philosophy of Statistical Theories in The Physical Sciences

1976, xii + 241 pp.
Cloth Dfl. 75.- / US \$29.00
Paper Dfl. 45.- / US \$17.50

7. Logic, Methodology and Philosophy of Science

Proceedings of the 5th International Congress
of Logic, Methodology and Philosophy of
Science, held at London (Ontario),
August 27-September 2, 1975.

Edited by R. Butts and J. Hintikka
Set of 2 volumes
1976 (in preparation)



D. Reidel Publishing Company

P.O. Box 17, Dordrecht-Holland

Lincoln Building, 160 Old Derby Street, Hingham, Mass. 02043 - U.S.A.

Volume LXXXIV

Number 336

MIND—OCTOBER 1976

Edited by D. W. HAMLYN

Articles:

Identity and Personal Identity
On What Sorts of Thing There Are
The Objectivity of Pain
Spinoza on the Mind-Body Problem: Two Questions
Wiggins on Identity

E. J. BOROWSKI
L. STEVENSON
N. FLEMING
C. B. DANIELS
H. W. NOONAN

Discussions:

Space and Time Re-assimilated
Sophist 263B Revisited
Negative Utilitarianism: Not Dead Yet
The Prescriptivism Incompleteness Theorem
The Identity of Laws: A Reply to Mr. Griffin
A Note on Belief

B. MAYO
K. M. SAYRE
R. I. SIKORA
H. J. GENSLE
R. HARRÉ
D. M. JOHNSON

Critical Notice:

The Central Questions of Philosophy by A. J. Ayer

J. HARRISON

Book Reviews

Books Received

Published for the MIND ASSOCIATION
by BASIL BLACKWELL

Single issues: £1.25

Annual Subscription: £4.00

The Monist *An International Quarterly Journal of General Philosophical Inquiry*

Founded 1888 by Edward C. Hegeler

Editor, Eugene FREEMAN

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor, Ann FREEMAN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate with the Editor in advance for Special Instructions. Papers should be 4000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.

SCHEDULED PUBLICATION DATES:

Vol. 60,	No. 1	Jan.,	1977	Bioethics and Social Responsibility
Vol. 60,	No. 2	Apr.,	1977	Philosophy and Religion in the Nineteenth Century
Vol. 60,	No. 3	July,	1977	New Directions in Semantics
Vol. 60,	No. 4	Oct.,	1977	Historicism and Epistemology
Vol. 61,	No. 1	Jan.,	1978	Conceptions of the Self: East and West
Vol. 61,	No. 2	Apr.,	1978	The Philosophy of Thomas Reid
Vol. 61,	No. 3	July,	1978	Nominalism, Past and Present
Vol. 61,	No. 4	Oct.,	1978	The Philosophy and Psychology of Cognition

Editorial Office: Box 1908, Los Gatos, California 95030

Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, \$18.00; Individuals, \$12.00. Single copies: \$4.50. Foreign postage: Add 25 cents to single copy rate or \$1.00 to subscription rate.

AMERICAN PHILOSOPHICAL QUARTERLY.

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

In March 1976 we published *NEGATION AND NON-BEING* by Richard Gale, which is the tenth in this impressive Monograph series. The price is \$8.00 (£4.00) to institutional subscribers and to non-subscribers; individual subscribers, however, will continue to receive the Monograph as part of their subscription. In past years we were able to supply Monographs to all subscribers to the Quarterly, but this has for the time being become impossible. If your library is a subscriber to the Quarterly, please check whether it is now subscribing also to the Monograph series. It is only in the Monographs that the larger themes of modern philosophy can receive extended treatment, and those who subscribe only to the Quarterly are missing an essential part of the coverage of philosophy which is provided by the journal as a whole.

To enable you to judge for yourself the importance of this series, a complete list of the Monographs is given overleaf. The last issue supplied to institutions on subscription was No. 7 (1973). If you find that Nos. 8 (1974) and 9 (1975) are still missing from your files, we shall be glad to supply them at the quoted prices, and to enter your standing order for all future issues of the series.

PRINTED IN ENGLAND

by William Clowes & Sons, Limited, London, Beccles and Colchester

AMERICAN PHILOSOPHICAL QUARTERLY MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

- No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielson, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$8.00 (£4.00).
- No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$8.00 (£4.00).
- No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$8.00 (£4.00).
- No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, \$8.00 (£4.00).
- No. 5. DESCARTES' PHILOSOPHY OF NATURE by James Collins. 1971, \$8.00 (£4.00).
- No. 6. STUDIES IN THE PHILOSOPHY OF MIND. *Contents:* Harold Brown, "Perception and Meaning"; Haskell Fain and A. P. Griffiths, "On Falsely Believing That One Doesn't Know"; O. H. Green, "Emotions and Belief"; Donald F. Henze, "Descartes on Other Minds"; Moreland Perkins, "Meaning and Feeling"; and Alan R. White, "What We Believe." 1972, \$8.00 (£4.00).
- No. 7. STUDIES IN ETHICS. *Contents:* Charles V. Blatz, "Mitigating and Meliorating Defenses"; Claudia Card, "Retributive Penal Liability"; Norman O. Dahl, "Is Mill's Hedonism Inconsistent?"; R. L. Franklin, "Recent Work on Ethical Naturalism"; James D. Wallace, "Cowardice and Courage"; A. D. Woozley, "Injustice." 1973, \$8.00 (£4.00).
- No. 8. STUDIES IN MODALITY by Nicholas Rescher, with the collaboration of Ruth Manor, Arnold vander Nat, and Zane Parks. 1974, \$8.00 (£4.00).
- No. 9. STUDIES IN EPISTEMOLOGY. *Contents:* Dennis M. Ahern, "Hume on the Evidential Impossibility of Miracles"; Robert Almeder, "Fallibilism and the Ultimate Irreversible Opinion"; Jonathan Bennet, "Stimulus, Response, Meaning"; Peter A. French, "'Seeing' and 'Seeing that', 'Observing' and 'Observing that'"; Margaret Gilbert and Fred R. Berger, "On an Argument for the Impossibility of Prediction in the Social Sciences"; David K. Glidden, "Protagorean Relativism and Cyrenaica"; Mark Pastin, "Modest Foundationalism and Self-Warrant"; E. E. Sleinis, "Normal Conditions and Perceivers." 1975, \$8.00 (£4.00).